The Structure of Inference

D. A. S. FRASER

Professor of Mathematics
University of Toronto

John Wiley & Sons, Inc.

New York. London. Sydney

WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

Probability and Mathematical Statistics

ALEXANDER · Elements of Mathematical Statistics
ANDERSON · An Introduction to Multivariate Statistical Analysis
BLACKWELL and GIRSHICK · Theory of Games and Statistical Decisions
CRAMÉR · The Elements of Probability Theory and Some of Its Applications
DOOB · Stochastic Processes
FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*
FELLER · An Introduction to Probability Theory and Its Applications, Volume II
FISHER · Contributions to Mathematical Statistics
FISZ · Probability Theory and Mathematical Statistics, *Third Edition*
FRASER · Nonparametric Methods in Statistics
FRASER · Statistics—An Introduction
FRASER · The Structure of Inference
GRENANDER and ROSENBLATT · Statistical Analysis of Stationary Time Series
HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory, Volume II
HOEL · Introduction to Mathematical Statistics, *Third Edition*
KEMPTHORNE · The Design and Analysis of Experiments
LEHMANN · Testing Statistical Hypotheses
PARZEN · Modern Probability Theory and Its Applications
RAO · Linear Statistical Inference and Its Applications
RIORDAN · An Introduction to Combinatorial Analysis
SCHEFFÉ · The Analysis of Variance
WALD · Sequential Analysis
WILKS · Collected Papers: Contributions to Mathematical Statistics
WILKS · Mathematical Statistics

Applied Probability and Statistics

BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
BENNETT and FRANKLIN · Statistical Analysis in Chemistry and the Chemical Industry
BROWNLEE · Statistical Theory and Methodology in Science and Engineering, *Second Edition*
BUSH and MOSTELLER · Stochastic Models for Learning
CHAKRAVARTI, LAHA and ROY · Handbook of Methods of Applied Statistics, Vol. I

Applied Probability and Statistics (Continued)

CHAKRAVARTI, LAHA and ROY · Handbook of Methods of Applied Statistics, Vol. II
CHERNOFF and MOSES · Elementary Decision Theory
CHEW · Experimental Designs in Industry
CHIANG · Introduction to Stochastic Processes in Biostatistics
CLELLAND, deCANI, BROWN, BURSK, and MURRAY · Basic Statistics with Business Applications
COCHRAN · Sampling Techniques, *Second Edition*
COCHRAN and COX · Experimental Designs, *Second Edition*
COX · Planning of Experiments
COX and MILLER · The Theory of Stochastic Processes
DEMING · Sample Design in Business Research
DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
DRAPER and SMITH · Applied Regression Analysis
GOLDBERGER · Econometric Theory
GUTTMAN and WILKS · Introductory Engineering Statistics
HALD · Statistical Tables and Formulas
HALD · Statistical Theory with Engineering Applications
HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory, Volume I
HOEL · Elementary Statistics, *Second Edition*
JOHNSON and LEONE · Statistics and Experimental Design: In Engineering and the Physical Sciences, Volumes I and II
KEMPTHORNE · An Introduction to Genetic Statistics
MEYER · Symposium on Monte Carlo Methods
PRABHU · Queues and Inventories: A Study of Their Basic Stochastic Processes
SARHAN and GREENBERG · Contributions to Order Statistics
TIPPETT · Technological Applications of Statistics
WILLIAMS · Regression Analysis
WOLD and JURÉEN · Demand Analysis
YOUDEN · Statistical Methods for Chemists

Tracts on Probability and Statistics

BILLINGSLEY · Ergodic Theory and Information
CRAMÉR and LEADBETTER · Stationary and Related Stochastic Processes
RIORDAN · Combinatorial Identities
TAKÁCS · Combinatorial Methods in the Theory of Stochastic Processes

The Structure of Inference

The Structure of Inference

D. A. S. FRASER

Professor of Mathematics
University of Toronto

John Wiley & Sons, Inc.
New York. London. Sydney

Library of Congress Catalog Card Number: 68-19333
Printed in the United States of America

Preface

Statistical inference is concerned with unknowns in scientific investigations —physical constants, properties, relationships—unknowns whose effects can be discovered in whole or in part by the manipulation of some variables and the observation and measurement of others. Statistical inference is the theory that describes and prescribes the argument from observation and measurement to conclusion about the unknowns. Envisaged broadly, it encompasses the scientific method. Interpreted in familiar form, it tends to be restricted to observation and design with the basic variables already chosen and to instances of application that have significant variation beyond that generated by the manipulated variables.

This book presents a unified theory of statistical inference. It is organized and developed as an introductory text on mathematical statistics. It presupposes a familiarity with elementary probability theory (a first course), with elementary vector analysis, and with multiple differentiation and integration (a second course in calculus).

Statistical inference requires a statistical model—a model that describes the essential aspects of the process or experiment being investigated. Most processes and experiments contain sources of variation—identifiable sources that can be described by means of error variables; for example, the error in the operation of a measuring instrument, the variation in the raw material to a process, the variation in the interactions within a process, and the variation due to the randomization component of an experimental design. In such cases the statistical model must include the appropriate error variable.

In many processes and experiments the observed response value is generated by a simple kind of transformation of a realized error value. The first three chapters examine models that have an error variable and the simple kind of transformation: in Chapter One models for the direct measurement of a physical quantity; in Chapter Two the general model; and in Chapter Three a range of models for the indirect measurement of physical quantities. The analysis and the inference apply equally to any error form and are not restricted to the traditional normal or Gaussian error distribution.

As a by-product, the first three chapters produce much of the standard distribution theory for the classical statistical model; the classical model

neglects the error variable and describes only the response variable of the process or experiment. The derivations do not employ the usual moment-generating functions and convolution formulas but rather an elementary device based on transformations. The method of derivation is simpler and applies to any error form, not just the normal.

The middle three chapters examine models that have the error variables but also quantities that are not in direct correspondence with the simple kind of transformation. New methods of analysis are obtained, such as the method of marginal likelihood. And exact solutions are obtained for problems inaccessible by the traditional methods; for example, data transformations for the regression model in Chapter Four.

As a by-product the middle three chapters produce the standard distribution-theory results of multivariate analysis. Again the results are obtained by the elementary device based on transformations and are available for any error form, not just the normal.

The last three chapters examine inference for the classical statistical model—for applications in which the error variable cannot be identified. For a large number of observations the error-variable structure arises in the accessible model and the methods in the preceding chapters are available.

The book contains material additional to a two-semester course in mathematical statistics. Appropriate sections for deletion on a first reading and more difficult problems are marked by an asterisk. Answers to selected problems are recorded in an appendix. A solution booklet is available to attested instructors from Y. S. Lee and the author, Department of Mathematics, University of Toronto.

The material in this book was developed over a ten-year period at the University of Toronto. The development was furthered by the opportunity to visit other universities and present and discuss various portions of the material: Stanford University, 1961–1962, the University of California, Berkeley, 1963, the University of Copenhagen, 1964, and the University of Wisconsin, 1965. The preparation of the final versions of the manuscript was made possible by support from the National Research Council of Canada.

I value very much the help and advice of friends: Geoffrey S. Watson who read the preliminary and final manuscripts and gave crucial advice; M. Safiul Haq and W. Keith Hastings who examined and discussed the preliminary manuscript; M. Masoom Ali, James Bondar, and Leonard Steinberg, who read the final manuscript in fine detail and broad pattern; Andrew Kalotay, Hans Levenbach, Leonard Steinberg, and Jim Whitney, who worked closely on the development of sections in Chapter Two and Four; Y. S. Lee who carefully checked and solved the problems; Bob Montgomery who prepared the drawings in perspective; and Iris Martin and Mary O'Rourke who carefully typed the manuscript.

DONALD FRASER

Contents

PART II METHODS OF ANALYSIS

PART III EXTENSIONS

The Structure of Inference

PART I

Foundations

CHAPTER ONE

Measurement Models

The eighteenth and nineteenth centuries saw the gradual emergence of a new discipline, a discipline eventually to receive the name *statistical inference*. This new discipline arose as part of probability theory, but its problems were different—both in kind and in purpose. A typical problem of probability theory concerned gambling games involving randomness, and the purpose was to derive models and calculate probabilities. A typical problem of statistical inference concerned measurement error, and the purpose was to infer the values of the quantities being measured. The purpose, *inference*, became the distinguishing feature of statistical inference, while the subject *measurement error*, receded from general attention. The neglect of measurement error cannot be attributed to success of the theory in treating the topic. Rather it indicates failure, the accumulated theory coming to partial agreement only for very special cases such as with normally distributed error.

This chapter considers two kinds of problem involving measurement error. It finds an essential ingredient for a measurement model, an ingredient effectively absent in the accumulated theory; and with this ingredient included it derives the general solution for inference, a solution applicable to any error form, normal or nonnormal. More general problems involving indirect measurement are examined in Chapter Three.

THE SIMPLE MEASUREMENT MODEL

1 THE MODEL

Consider an instrument I for measuring a certain kind of physical quantity. Suppose the operation of I has been investigated; and suppose its error pattern in repetitions has been found describable as independent realizations of an *error variable* e with probability element $f(e)\,de$ on the real line R^1 (see Figure 1). A value of the error variable gives the difference between a reading of the instrument and a value of the quantity.

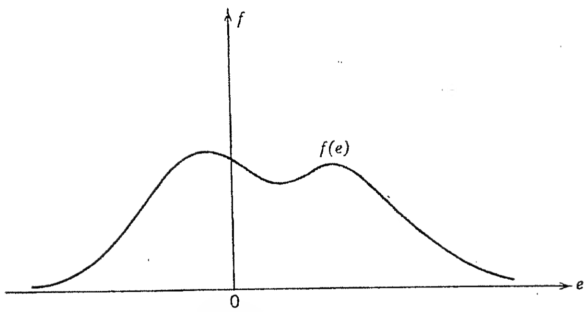


Figure 1 The error distribution of an instrument.

Now consider the use of I for a single measurement on a quantity. Let x be the *value* of the measurement and θ be the *value* of the quantity. The operation of the instrument and the instance of measurement can then be described by the model

$$f(e)\, de,$$
$$x = \theta + e.$$

The model has two parts: an *error distribution* $f(e)\, de$ which describes the operation of the instrument (with e as a *variable*) and a *structural equation* $x = \theta + e$ in which a *realized* value e from the error distribution has determined the relation between the value x of the measurement and the value θ of the quantity (with e as a *constant*). This is illustrated schematically in Figure 2.

Now consider the use of the instrument for multiple measurements on a physical quantity. The multiple operation of the instrument in a sequence of n operations has probability element $\Pi f(e_i)\, \Pi\, de_i$ on Euclidean space R^n. Let $(x_1, \ldots, x_n) = \mathbf{x}'$ be the sequence of measurement values and θ be the value of the quantity. The operation of the instrument and the n instances of measurement can then be described by the

Simple Measurement Model

$$\prod_1^n f(e_i) \prod_1^n de_i,$$
$$x_1 = \theta + e_1,$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$x_n = \theta + e_n.$$

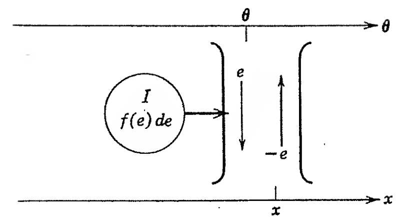


Figure 2 The simple measurement model, $n = 1$.

The model has two parts: an *error distribution* $\Pi f(e_i)\, \Pi\, de_i$ which describes the multiple operation of the measuring instrument (with e's as *variables*), and a *structural equation* $\mathbf{x} = \theta\mathbf{1} + \mathbf{e}$ (in vector notation) in which a *realized* vector $\mathbf{e}$ from the error distribution has determined the relation between the known measurement vector $\mathbf{x}$ and the unknown value θ of the quantity (with $\mathbf{e}$ as a *constant*).

2 THE TRANSFORMATIONS

In the structural equation $x = \theta + e$ a realized error value e is translated by an amount θ. In the modified equation $\theta = x - e$, a reverse error value $-e$ is translated by an amount x. Translations such as these are integral to the use of the instrument.

Consider notation for translations, notation general enough to cover the rescalings of interest in later sections. Let $[a, c]$ be the *affine transformation* $(c \neq 0)$ on R^1,

$$[a, c]x = a + cx,$$

or the corresponding affine transformation on R^n,

$$[a, c]\mathbf{x} = a\mathbf{1} + c\mathbf{x}.$$

The *composition* or *product* of two affine transformations is affine:

$$[A, C][a, c]x = A + Ca + Ccx,$$
$$[A, C][a, c] = [A + Ca, Cc].$$

Note that the product depends on the order of the component transformations. The *identity* transformation is affine:

$$[0, 1][a, c] = [a, c] = [a, c][0, 1].$$

And the *inverse* of an affine transformation is affine:

$$[a, c]^{-1} = [-c^{-1}a, c^{-1}],$$
$$[-c^{-1}a, c^{-1}][a, c] = [0, 1] = [a, c][-c^{-1}a, c^{-1}].$$

A set of transformations that is closed under the formation of products and inverses has the algebraic structure of a group:

Definition 1. A set G is a group if (i) for each pair (g_1, g_2) of elements of G, there is an element g_1g_2 of G called the product of g_1 and g_2; (ii) for each triple (g_1, g_2, g_3) of elements of G, $(g_1g_2)g_3 = g_1(g_2g_3)$ (associativity); (iii) there is an element i of G called the identity with the property $ig = gi = g$ for each element g of G; (iv) for each element g of G there is an element g^{-1} of G with the property $gg^{-1} = g^{-1}g = i$.

The affine transformations

$$\{[a, c]: \quad -\infty < a < \infty, \quad c \neq 0\}$$

form a group, the *affine group on* R^1.

The translations on R^1 or the corresponding translations on R^n have the form

$$[a, 1]x = a + x, \qquad [a, 1]\mathbf{x} = a\mathbf{1} + \mathbf{x}.$$

These transformations form the *location group on* R^1:

$$G = \{[a, 1]: \quad -\infty < a < \infty\};$$

the group properties are

$$[A, 1][a, 1] = [A + a, 1],$$
$$[a, 1]^{-1} = [-a, 1], \qquad i = [0, 1].$$

The *simple measurement model* can now be re-expressed by using the transformation notation:

$$\prod_1^n f(e_i) \prod_1^n de_i,$$
$$\mathbf{x} = [\theta, 1]\mathbf{e}.$$

3 THE ORBITS

Consider how the location group G affects Euclidean space R^n. The translations $[a, 1]$ carry a point $\mathbf{x}$ into the points $a\mathbf{1} + \mathbf{x}$ (see Figure 3). These points form the *orbit of* $\mathbf{x}$ under the location group:

$$G\mathbf{x} = \{[a, 1]: \quad -\infty < a < \infty\}\mathbf{x} = \{a\mathbf{1} + \mathbf{x}: \quad -\infty < a < \infty\}.$$

The orbit that passes through the origin $\mathbf{0}$ has the special form

$$G\mathbf{0} = \{a\mathbf{1}: \quad -\infty < a < \infty\};$$

it is a *one-dimensional linear subspace*, the *extended* $\mathbf{1}$*-vector*. The general orbit $G\mathbf{x}$ can be formed by an $\mathbf{x}$-translation of $G\mathbf{0}$:

$$G\mathbf{x} = G\mathbf{0} + \mathbf{x} = \{a\mathbf{1} + \mathbf{x}\};$$

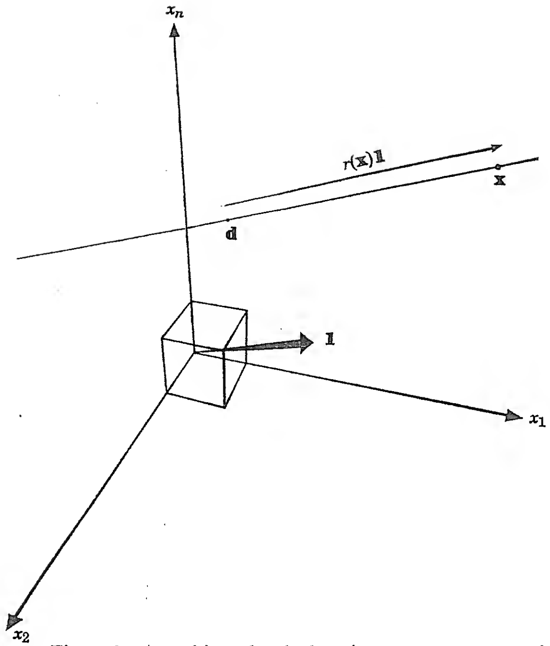


Figure 3 An orbit under the location group.

it is a *one-dimensional affine subspace*. Clearly, two orbits are either identical or disjoint.

The effect of the location group can be examined alternatively by treating $(x_1, \ldots, x_n)$ as n numbered points on the real line. A translation $[a, 1]$ carries $x_1, \ldots, x_n$ into $\tilde{x}_1, \ldots, \tilde{x}_n$, where $\tilde{x}_i = a + x_i$. The order of the points on the real line is unchanged; the spacings between the points are unchanged. See Figure 4. Only the location of the n points is affected by the translation.

Consider a variable to describe the position of a point $\mathbf{x}$ on its orbit.

Definition 2. $r(\mathbf{x})$ is a location variable if

$$r([a, 1]\mathbf{x}) = [a, 1]r(\mathbf{x})$$

for all $\mathbf{x}$ and a; that is, *if*

$$r(a\mathbf{1} + \mathbf{x}) = a + r(\mathbf{x})$$

for all $\mathbf{x}$ and a.

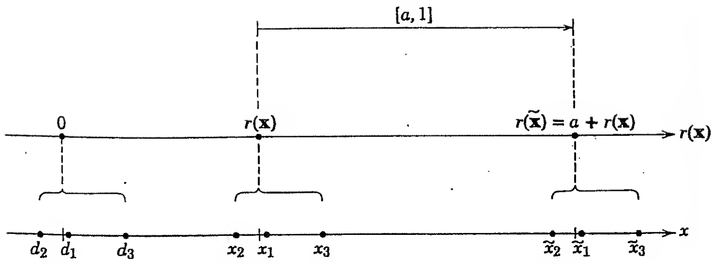


Figure 4 A translation $[a, 1]$ of three points.

As examples consider $\bar{x} = \Sigma\, x_i/n$, $\max x_i$, x_1, $(\max x_i + \min x_i)/2$, $2x_3 - x_1$, $6 \max x_i - 5 \min x_i$ (see Figure 4). A location variable is linear in a restricted sense: linear along any orbit.

A location variable $r(\mathbf{x})$ leads to a *reference point* $\mathbf{d}(\mathbf{x})$ on each orbit, *the point at which the variable takes the value* 0:

$$\begin{aligned}
\mathbf{d}(\mathbf{x}) &= [r(\mathbf{x}), 1]^{-1}\mathbf{x} \\
&= -r(\mathbf{x})\mathbf{1} + \mathbf{x} \\
&= (x_1 - r(\mathbf{x}), \ldots, x_n - r(\mathbf{x}))'; \\
r(\mathbf{d}) &= r(-r(\mathbf{x})\mathbf{1} + \mathbf{x}) \\
&= -r(\mathbf{x}) + r(\mathbf{x}) = 0.
\end{aligned}$$

The reference points $\mathbf{d}(\mathbf{x})$ index the orbits $G\mathbf{x}$ (see Figures 3 and 4); each orbit has exactly one reference point.

Each point $a\mathbf{1} + \mathbf{d}$ on the orbit through a reference point $\mathbf{d}$ has a different *position*:

$$r(a\mathbf{1} + \mathbf{d}) = a + r(\mathbf{d}) = a.$$

Note that $r(\mathbf{x})$ measures position using the reference point as *origin* and the $\mathbf{1}$-vector as *unit*.

The general point $\mathbf{x}$ can be reconstructed from its orbit and its position:

$$\mathbf{x} = [r(\mathbf{x}), 1]\mathbf{d}(\mathbf{x});$$

for example, with $r(\mathbf{x}) = x_1$:

$$\begin{aligned}
\mathbf{d}(\mathbf{x}) &= (d_1(\mathbf{x}), \ldots, d_n(\mathbf{x}))' \\
&= (0, x_2 - x_1, \ldots, x_n - x_1)'; \\
\mathbf{x} &= [x_1, 1](0, x_2 - x_1, \ldots, x_n - x_1)'.
\end{aligned}$$

Two location variables differ in value by a constant along any orbit:

$$r_2(a\mathbf{1} + \mathbf{x}) - r_1(a\mathbf{1} + \mathbf{x}) = r_2(\mathbf{x}) - r_1(\mathbf{x}) = r_2(\mathbf{d}_1(\mathbf{x})).$$

Now consider the simple measurement model:

$$\prod f(e_i) \prod de_i,$$
$$\mathbf{x} = [\theta, 1]\mathbf{e}.$$

And let $r(\mathbf{x})$ be a location variable. The points $\mathbf{x}$ and $\mathbf{e}$ are on the same orbit:

$$G\mathbf{x} = G[\theta, 1]\mathbf{e} = G\mathbf{e} \quad \text{or} \quad \mathbf{d}(\mathbf{x}) = \mathbf{d}(\mathbf{e}).$$

The *positions* of the points $\mathbf{x}$ and $\mathbf{e}$ differ by a translation $[\theta, 1]$:

$$r(\mathbf{x}) = [\theta, 1]r(\mathbf{e}).$$

The simple measurement model can then be rewritten with a composite structural equation:

$$\prod f(e_i) \prod de_i,$$
$$r(\mathbf{x}) = [\theta, 1]r(\mathbf{e}), \qquad G\mathbf{x} = G\mathbf{e}.$$

4 HOMOGENEITY

Consider a transformation $[a, 1]$ on the axis of measurement,

$$\tilde{\mathbf{x}} = [a, 1]\mathbf{x}, \qquad \tilde{\theta} = [a, 1]\theta;$$

and view the transformation as providing *new* coordinates for *given* points. See Figure 5. The transformation does not affect the physical problem of measurement; it affects only the numerical representation of the values involved.

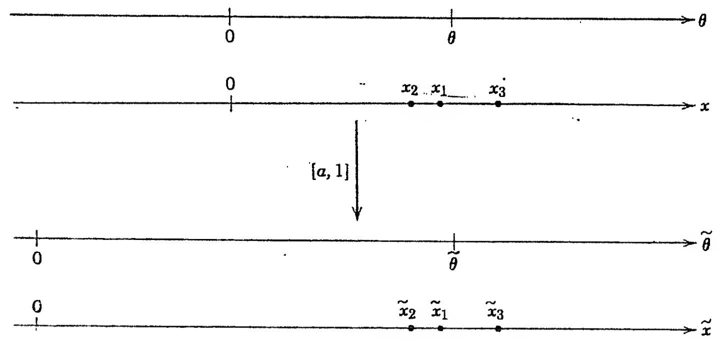


Figure 5 A change of coordinates $[a, 1]$.

Consider how the change of coordinates affects the simple measurement model:

$$\prod f(e_i) \prod de_i,$$
$$\mathbf{x} = [\theta, 1]\mathbf{e}.$$

The structural equation can be expressed in terms of the new coordinates: multiply by $[a, 1]$ and simplify using $[a, 1][\theta, 1] = [\tilde{\theta}, 1]$. The equation becomes

$$\tilde{\mathbf{x}} = [\tilde{\theta}, 1]\mathbf{e}.$$

Thus the model as expressed in terms of the new coordinates is

$$\prod f(e_i) \prod de_i,$$
$$\tilde{\mathbf{x}} = [\tilde{\theta}, 1]\mathbf{e}.$$

The form of the model is the same as before the change of coordinates. The physical problem is untouched by a transformation $[a, 1]$. Reflecting this, the model has the *same form* after the transformation as before. The simple measurement model is *homogeneous under the location group*.

The homogeneity can be pictured in terms of the axis of measurement: the measurement model as viewed from one point on the axis has the same form as when viewed from any other point on the axis.

5 PROBABILITIES FOR AN UNKNOWN CONSTANT

In applications of probability theory it is common to make probability statements concerning unknown constants. Consider briefly the conditions for such statements.

As an illustration suppose a deck of 52 playing cards is thoroughly shuffled and two cards are dealt face down on a table. The designations on the faces of the two cards are fixed; the designations are *unknown constants*. An observer can make probability statements concerning the unknown constants; for example,

$$\Pr\{2 \text{ spades}\} = \tfrac{13}{52} \cdot \tfrac{12}{51}.$$

Such statements are based on the *random process that generated the unknown constants*.

Now suppose two more cards are dealt from the deck face down on the table, and suppose the observer examines these cards and finds the first to be a spade and the second a nonspade. The observer can then make revised probability statements concerning the unknown constants; for example,

$$\Pr\{2 \text{ spades}\} = \frac{\tfrac{13}{52} \cdot \tfrac{12}{51} \cdot \tfrac{11}{50} \cdot \tfrac{39}{49}}{\tfrac{13}{52} \cdot \tfrac{39}{51}} = \tfrac{13}{50} \cdot \tfrac{12}{49} \cdot \tfrac{11}{13}.$$

Such statements are based on the *random process as conditioned by the observed event*.

Alternatively, suppose the second pair of cards is kept face down and passed to a participant in an adjacent room, and suppose the participant reports the item of information, "There's a spade here." The observer might then make the statement

$$\Pr\{2 \text{ spades}\} = \frac{\tfrac{13}{52} \cdot \tfrac{12}{51} \cdot \tfrac{11}{50}}{\tfrac{13}{52}} = \tfrac{13}{51} \cdot \tfrac{12}{50} \cdot \tfrac{11}{13},$$

if he thought the participant had examined only the first card. Or he might make the statement

$$\Pr\{2 \text{ spades}\} = \frac{\tfrac{13}{52} \cdot \tfrac{12}{51}(2 \cdot \tfrac{11}{50} \cdot \tfrac{39}{49})}{2 \cdot \tfrac{13}{52} \cdot \tfrac{39}{51}} = \tfrac{13}{50} \cdot \tfrac{12}{49} \cdot \tfrac{11}{13},$$

if he thought the participant had examined both cards and would have reported two spades if there were two spades. The two statements are contradictory.

For this alternative situation an exact probability statement for the value of Pr {2 spades} cannot be made. The item of information, "There's a spade here," *could* have been presented for each possible second pair having one or more spades. For exact probability statements it is necessary to know exactly those second pairs for which the item of information *would* have been presented. Information needs to be in the form of an *event*, the set of possible outcomes for which the information *would have been presented*. The item of information, "There's a spade here," has the form of a *deduction* from an event unknown to the observer.

The example illustrates sufficient conditions for making probability statements concerning unknown constants: (i) *The constants were generated as realized values from a random process with known probability characteristics.* (ii) *The only other information concerning the unknown constants has the form of an event for the random process that generated the constants.*

6 REDUCTION

Consider the measurement of a physical quantity. Let $x_1, \ldots, x_n$ be the measurements and θ the value of the quantity; and suppose the simple measurement model is applicable:

$$\prod f(e_i) \prod de_i,$$
$$r(\mathbf{x}) = [\theta, 1]r(\mathbf{e}), \qquad G\mathbf{x} = G\mathbf{e}.$$

The error distribution $\prod f(e_i) \prod de_i$ on R^n describes the operation of the measuring instrument; it describes the *random process* that generated the realized errors $e_1, \ldots, e_n$ in the structural equation. The structural equation in composite form gives the relation between the *known* values $x_1, \ldots, x_n$ and the *unknown* values $\theta, e_1, \ldots, e_n$.

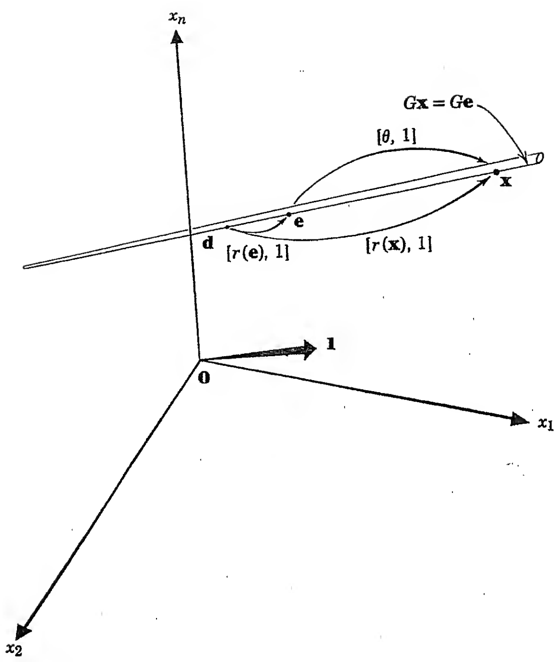


Figure 6 The known orbit of e and an adjacent bundle of orbits.

Suppose there is no other information concerning the unknowns; this can occur minimally if the measurement process is being examined in isolation to determine what information it alone supplies about the value θ. Now consider the information in the structural equation concerning the unknown error vector **e**.

The orbit of **e** is known: $G\mathbf{e} = G\mathbf{x}$. Or if **x** is known only to a certain accuracy, then the orbit of **e** is one of a bundle of orbits, as indicated in Figure 6. The information about the orbit of **e** is in *the form of an event for the process that generated* **e**, an event based on the partition of R^n into orbits.

Consider the position of **e** on its orbit. The position part of the structural equation can be solved:

$$r(\mathbf{e}) = [\theta, 1]^{-1} r(\mathbf{x}).$$

The position of **e** is described as an unknown translation, $g = [\theta, 1]^{-1}$, of the known position $r(\mathbf{x})$:

$$r(\mathbf{e}) = g r(\mathbf{x});$$

the error position is not known. If the known position $r(\mathbf{x})$ were different, $[a, 1]r(\mathbf{x})$ for example, then the structural equation would describe $r(\mathbf{e})$ as

$$r(\mathbf{e}) = g[a, 1]r(\mathbf{x}) = hr(\mathbf{x}),$$

where $h = g[a, 1]$ is also an unknown translation (homogeneity of the model). Different values of the position $r(\mathbf{x})$ would provide the same description for $r(\mathbf{e})$. There is thus no information from the structural equation concerning the location of **e** on its orbit.

IN SUMMARY. The error distribution describes the random process that generated the unknown **e** in the structural equation. The only other information concerning the unknown **e** has the form $G\mathbf{e} = G\mathbf{x}$, an event for the random process that generated **e**. The conditions are fulfilled—*exact probability statements can be made concerning the unknown error* **e**; they are based on the conditional distribution of the error variable **e** given the orbit $G\mathbf{e} = G\mathbf{x}$.

7 THE REDUCED MODEL

Consider the derivation of the conditional distribution of the error variable **e** given the orbit $G\mathbf{e}$. On any orbit, two location variables differ in value by a constant. It suffices to work with a simple choice; take $r(\mathbf{e}) = e_1$. The corresponding **d**-vector has coordinates

$$\begin{aligned} d_1(\mathbf{e}) &= e_1 - e_1 = 0, \\ d_2(\mathbf{e}) &= e_2 - e_1, \\ &\vdots \\ d_n(\mathbf{e}) &= e_n - e_1. \end{aligned}$$

The required conditional distribution is then the distribution of e_1 given $d_2, \ldots, d_n$.

The probability element for **e** is

$$\prod_1^n f(e_i) \prod_1^n de_i.$$

The Jacobian determinant of $(e_1, d_2, \ldots, d_n)$ with respect to $(e_1, \ldots, e_n)$ is

$$J = \begin{vmatrix} 1 & & & & 0 \\ -1 & 1 & & & \\ -1 & 0 & 1 & & \\ \vdots & & & \ddots & \\ -1 & 0 & \cdots & 0 & 1 \end{vmatrix} = 1;$$

hence the probability element for $(e_1, d_2, \ldots, d_n)$ is

$$\prod_1^n f(e_1 + d_i)\, de_1\, dd_2 \cdots dd_n.$$

The marginal probability element for $(d_2, \ldots, d_n)$ is

$$h(d_2, \ldots, d_n)\, dd_2 \cdots dd_n = \int_{-\infty}^{\infty} \prod_1^n f(t + d_i)\, dt \cdot dd_2 \cdots dd_n;$$

hence the conditional probability element for e_1 given $d_2, \ldots, d_n$ is

$$g(e_1 : \mathbf{d})\, de_1 = \frac{\prod_1^n f(e_1 + d_i)}{h(d_2, \ldots, d_n)}\, de_1 = \frac{\prod_1^n f(e_1 + d_i)}{\int_{-\infty}^{\infty} \prod_1^n f(t + d_i)\, dt}\, de_1$$

$$= k(\mathbf{d}) \prod_1^n f(e_1 + d_i)\, de_1.$$

The denominator in the middle two expressions serves only as a normalizing constant.

Let $r^* = r^*(\mathbf{e})$ be an alternative location variable:

$$e_i = e_1 + d_i = r^* + d_i^*.$$

The conditional probability element can be reexpressed in terms of the new variable:

$$k^*(\mathbf{d}^*) \prod_1^n f(r^* + d_i^*)\, dr^*.$$

The distribution has the same form as before; and the normalizing constant has the same value—but is expressed in terms of the new reference point.

The conditional distribution described in the preceding section has now been derived. The simple measurement model by its own information content produces the

Reduced Simple Measurement Model

$$g(r : \mathbf{d}(\mathbf{x}))\, dr,$$
$$r(\mathbf{x}) = \theta + r.$$

The reduced model has two parts: an *error probability distribution* $g(r : \mathbf{d}(\mathbf{x}))\, dr$ on R^1 (with r as a *variable*) which provides probability statements for the unknown error position r in the structural equation; and a *structural equation* which gives the relation between the known $r(\mathbf{x})$ and the unknowns θ and r (with r as a *constant*).

8 EXAMPLES

As a first example consider the simple measurement model with error variable normally distributed with mean 0 and variance σ_0^2:

$$\prod f(e_i) \prod de_i = (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2} \sum e_i^2\right\} \prod de_i,$$
$$\mathbf{x} = \theta \mathbf{1} + \mathbf{e}.$$

The location variable $\bar{x}$ is particularly suited to the case of normal error. The conditional error distribution of $\bar{e}$ given $\mathbf{d}(\mathbf{e}) = (e_1 - \bar{e}, \ldots, e_n - \bar{e})' = \mathbf{d}$ is

$$g(\bar{e})\, d\bar{e} = k'(2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2} \sum (\bar{e} + d_i)^2\right\} d\bar{e}$$

$$= k'' \exp\left\{-\frac{n}{2\sigma_0^2} \bar{e}^2\right\} d\bar{e}$$

$$= \left(2\pi \frac{\sigma_0^2}{n}\right)^{-1/2} \exp\left\{-\frac{n}{2\sigma_0^2} \bar{e}^2\right\} d\bar{e}.$$

The first step in the simplification uses

$$\sum (\bar{e} + d_i)^2 = n\bar{e}^2 + 2\bar{e} \sum d_i + \sum d_i^2$$
$$= n\bar{e}^2 + \sum d_i^2,$$

and incorporates the contribution from $\Sigma\, d_i^2$ into the constant k''; the second step supplies the necessary normalizing constant for the normal distribution. The reduced model is thus

$$g(\bar{e})\, d\bar{e} = \left(2\pi \frac{\sigma_0^2}{n}\right)^{-1/2} \exp\left\{-\frac{n}{2\sigma_0^2} \bar{e}^2\right\} d\bar{e},$$
$$\bar{x} = \theta + \bar{e};$$

this can be expressed equivalently as

$$\bar{e} = \frac{\sigma_0}{\sqrt{n}} z,$$
$$\bar{x} = \theta + \bar{e},$$

where z designates a standard normal variable. The error distribution for the location variable $\bar{e}$ has the interesting property that it does not depend on the values of the deviations d_i; the distribution is the *same* on each orbit.

Suppose that $\sigma_0 = 0.6$; and suppose the measurements are

62.0, 60.5, 60.7, 61.6.

The simple measurement model is

$$e_1 = 0.6z_1, \qquad e_2 = 0.6z_2, \qquad e_3 = 0.6z_3, \qquad e_4 = 0.6z_4,$$
$$62.0 = \theta + e_1,$$
$$60.5 = \theta + e_2,$$
$$60.7 = \theta + e_3,$$
$$61.6 = \theta + e_4;$$

and the reduced model is

$$\bar{e} = 0.3z,$$
$$61.2 = \theta + \bar{e}.$$

The probability distribution describing the unknown $\bar{e}$ is normal with mean 0 and standard deviation 0.3. Some probability statements concerning $\bar{e}$ are

$$\Pr\{-0.3 \le \bar{e} \le 0.3\} = 68\tfrac{1}{4}\%,$$
$$\Pr\{-0.6 \le \bar{e} \le 0.6\} = 95\tfrac{1}{2}\%.$$

As a second example consider the simple measurement model with an error variable that has a Cauchy distribution in standard form; and suppose there are two measurements, 165.1, 161.1, on the quantity θ:

$$\prod f(e_i) \prod de_i = \frac{1}{\pi^2} \frac{1}{1 + e_1^2} \frac{1}{1 + e_2^2} de_1\, de_2,$$
$$165.1 = \theta + e_1,$$
$$161.1 = \theta + e_2.$$

The location variable x_1 is as convenient as any; the corresponding **d**-vector is $\mathbf{d} = (0, -4.0)'$. The reduced model is

$$g(e_1)\, de_1 = k \frac{1}{1 + e_1^2} \frac{1}{1 + (e_1 - 4)^2} de_1,$$
$$165.1 = \theta + e_1.$$

The conditional error distribution is plotted in Figure 7. The constant can be obtained by numerical integration from the graph itself. Probability statements concerning the unknown e_1 can also be derived from the graph; for example,

$$\Pr\{-1 \le e_1 \le 5\} \doteq 89.1\%$$

The simple measurement model has been developed with the measurement process as illustration. The range of applications, however, is much broader.

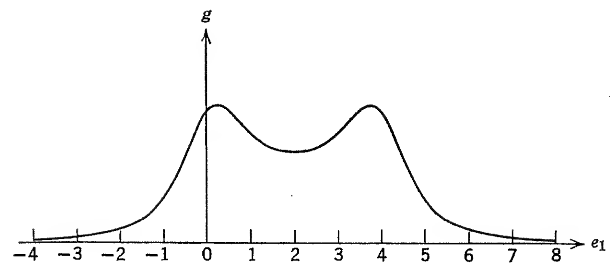


Figure 7 The error probability distribution for the Cauchy example.

A typical application has a sequence of observations or measurements on a response variable. The response variable is based on a process or system operating under stable conditions. Separate operations of the system are statistically independent as a consequence of separation in time, or separation in space, or separation in entity. The internal pattern of variation or *error* as it affects the response variable is known from earlier experience, and the spread of this error pattern is *also* known (or a more general model in Section 11 is needed). These characteristics of the typical application require a scale of measurement, and they *also* require a unit of measurement; they do not require an origin or zero point on the measurement scale.

Variation in a response variable can generally be attributed to a variety of sources: variation in the material being processed, variation in the internal operation of the process, variation due to the randomization ingredient of experimental design. The combined sources of variation form the *internal error* of the system; the composite effect of this error produces the variation or error that affects the response. The typical application requires that external conditions of the process be controlled and that sequencing of observations be randomized against possible external sources of variation. This procedure can provide the basis on which the internal error of the system has stability and the composite effect of this error has known form. The internal error of the system is the *random process* referred to in the development of the simple measurement model; and the composite effect of this error is the *error variable* described by the model.

In the typical application the medial or general level of the response is the quantity being investigated or *measured*. This quantity can have numerical definition by comparison with standard levels for similar variables. A zero point on the measurement scale may be chosen for convenience. In a typical process the general response level depends on the levels of input variables to the process. For the applications considered here, these input levels are kept constant and are part of the conditions of the system (more general models

that describe patterns of dependence on input variables are available in Chapter Three). The general response level is a consequence of the chosen conditions of the system; it has an identity distinct from the internal error of the system. The general response level is the *quantity* θ in the simple measurement model. The quantity θ gives response expression to the internal error values $e_1, \ldots, e_n$. The response expression is in the form of response *observations* or *determinations* $x_1, \ldots, x_n$, the *measurements* in the simple measurement model.

9 TESTS OF SIGNIFICANCE

In an application of the simple measurement model an outside source may indicate a value θ_0 for the quantity being measured. The outside source could be a preceding investigation on a similar quantity; the purpose of the application would then be to see whether the value of the quantity is the value indicated by that earlier investigation.

Alternatively, the outside source could be a theory linking a variety of physical quantities. The theory, perhaps in conjunction with values for some of the physical quantities, may prescribe a value θ_0 for the quantity being measured. The purpose of the application would then be to see whether the quantity being measured has the value θ_0, to check thereby whether the theory is adequate for the particular kind of prediction, and to check thereby on the validity of the theory.

As an illustration consider the first example in the preceding section. The reduced model is

$$\bar{e} = 0.3z,$$

$$61.2 = \theta + \bar{e},$$

where z designates a standard normal variable. Suppose that some outside source has prescribed the value $\theta_0 = 62.4$ for the quantity θ. The *hypothesis* $\theta = 62.4$ leads to the value

$$\bar{e} = 61.2 - 62.4 = -1.2 = -4(0.3)$$

for the error position $\bar{e}$. This value for $\bar{e}$ is -4 standard deviations from the center of the normal error distribution. The probability of a value so far or farther from the center of the distribution is extremely small:

$$\Pr\{|\bar{e} - 0| \geq 1.2\} = \Pr\{|z| \geq 4\} = 0.000{,}064.$$

The value -1.2 for error position is thus almost inconsistent with the error probability distribution; *in the framework of the model* it suggests strongly

that the hypothesis is not true, and in turn that a theory that produced the hypothesis is not true. An additional sequence of measurements on θ might strengthen or reverse the assessment.

Now consider the reduced simple measurement model

$$g(r)\,dr,$$

$$r(\mathrm{x}) = \theta + r,$$

and let θ_0 be a value prescribed for θ by some outside source. The *hypothesis* $\theta = \theta_0$ leads to the value

$$r = -\theta_0 + r(\mathrm{x})$$

for the error position r. This value for r can be compared with the probability distribution $g(r)\,dr$ for error position. A value in a broad central range of the distribution is a value *in accord with* the error distribution: the measurements are in accord or agreement with the hypothesis. A value in the extremes of the distribution is an unlikely value for the error distribution. Its *significance* can be assessed in part, as in the example, by calculating the *level of significance:* the probability of as great or greater departure from the center of the distribution. *Within the framework of the model*, an extreme value provides evidence against the hypothesis, and a value effectively beyond the range of the distribution effectively denies the hypothesis.

10 GENERAL INFERENCE

A primary need for statistical inference is the ability to extract information concerning an unknown quantity. A model, as it describes a system being investigated, contains the information about that system. Sometimes outside sources may also provide information. These two kinds of information *should* in general *be kept separate*, any combining being left to judgment and expediency on occasions when the information is used. Consider the simple measurement model and the information it contains concerning the unknown quantity.

Consider first the example at the beginning of Section 8. The reduced model is

$$g(\bar{e})\,d\bar{e} = \frac{1}{\sqrt{2\pi}\,0.3}\exp\left\{-\frac{1}{2(0.3)^2}\bar{e}^2\right\}d\bar{e},$$

$$61.2 = \theta + \bar{e}.$$

The reduced model contains all the information concerning unknown values. The error probability distribution provides probability statements concerning the unknown $\bar{e}$, and the structural equation links the unknowns $\bar{e}$ and θ.

Each possible value for $\bar{e}$ corresponds to a possible value for θ:

$$\theta = 61.2 - \bar{e}.$$

A probability statement concerning $\bar{e}$ is *ipso facto* a probability statement concerning θ. The probability statements concerning $\bar{e}$ are summarized in the distribution $g(\bar{e})\, d\bar{e}$; and correspondingly the probability statements concerning $\theta = 61.2 - \bar{e}$ are summarized in the distribution

$$g(61.2 - \theta)\, d\theta = \frac{1}{\sqrt{2\pi}\, 0.3} \exp\left\{-\frac{1}{2(0.3)^2}(\theta - 61.2)^2\right\} d\theta,$$

the *structural distribution* for the unknown value of θ. The structural distribution can be represented alternatively as

$$\theta = 61.2 - 0.3z,$$

where z is a standard normal variable. Some probability statements are

$$\Pr\{60.9 \le \theta \le 61.5\} = 68\tfrac{1}{4}\%,$$
$$\Pr\{60.6 \le \theta \le 61.8\} = 95\tfrac{1}{2}\%.$$

Now, more generally, consider the simple measurement model with normal error. The reduced model is

$$\bar{e} = \frac{\sigma_0}{\sqrt{n}} z,$$
$$\bar{x} = \theta + \bar{e}.$$

The structural distribution describing the unknown θ is given by

$$\theta = \bar{x} - \frac{\sigma_0}{\sqrt{n}} z.$$

Now, in general, consider the simple measurement model. The reduced model is

$$g(r)\, dr,$$
$$r(\mathbf{x}) = \theta + r.$$

The error probability distribution provides probability statements concerning the unknown r, and the structural equation links the two unknowns θ and r. Each possible value for r corresponds to a possible value for θ:

$$\theta = r(\mathbf{x}) - r.$$

The error distribution describing the unknown r is thus equivalent to the *structural distribution*

$$g(r(\mathbf{x}) - \theta)\, d\theta$$

describing the unknown value θ of the quantity.

THE MEASUREMENT MODEL

11 THE MODEL

Consider a system that can be operated under stable conditions. Suppose that, when operated repetitively under stable conditions, the internal error mechanisms produce a known error pattern—in a particular response as measured on a certain scale. And suppose this error pattern in some arbitrary units has the form of independent realizations of an *error variable* e with probability element $f(e)\, de$ on the real line R^1.

Now consider a particular set of conditions for the system. These conditions determine the characteristics of the response: the *spread* of the error pattern as given by a scale factor applied to the error variable (for numerical expression this requires a unit of measurement); and the *general level* of the response as given by a translation of the scaled error (for expression this needs an origin of measurement). Let $(x_1, \ldots, x_n) = \mathbf{x}'$ be a sequence of n observations on the response, and let σ be the unknown scale factor for the error and μ be the unknown general level of the response. The assumptions then give the

Measurement Model

$$\prod_1^n f(e_i) \prod_1^n de_i,$$
$$x_1 = \mu + \sigma e_1,$$
$$\vdots$$
$$x_n = \mu + \sigma e_n.$$

The model has two parts: an *error distribution* $\Pi f(e_i)\, \Pi\, de_i$ which describes the variation in the multiple operation of the system (with e's as *variables*); and a *structural equation* $\mathbf{x} = [\mu, \sigma]\mathbf{e}$ (vector notation) in which a realized vector $\mathbf{e}$ from the error distribution has determined the relation between the known observation $\mathbf{x}$ and the unknown system characteristics σ and μ (with $\mathbf{e}$ as a *constant*).

The conditions of the system determine the characteristics σ and μ; these characteristics appear as a transformation $[\mu, \sigma]$ which has a positive scaling factor σ and a relocation μ. Such a transformation is an element of the *positive affine group*

$$G = \{[a, c]:\ -\infty < a < \infty,\ \ 0 < c < \infty\}.$$

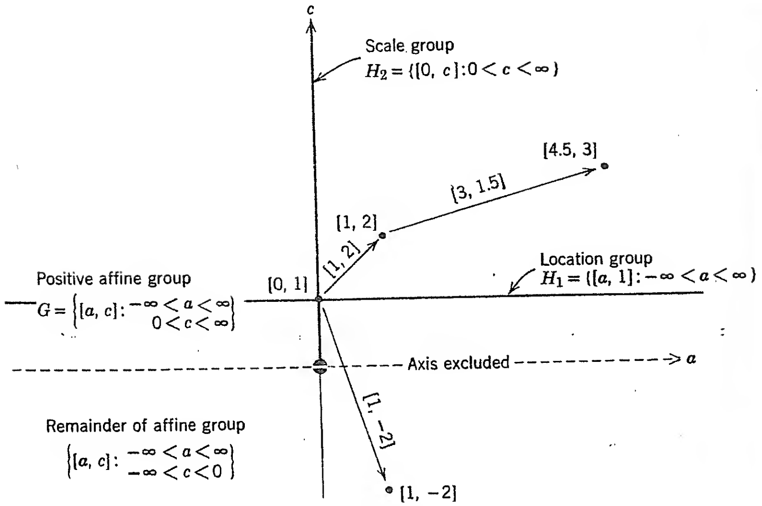


Figure 8 The affine group with three subgroups labeled. A group element $[a, c]$ can be represented as the *point* $[a, c]$ or as the *transformation* carrying $i = [0, 1]$ into the point $[a, c]$.

The set G is closed under the formation of products and inverses (formulas for product and inverse in Section 2). Accordingly G is a group.

A subset of a group that is itself a group using the same multiplication is called a *subgroup*. The positive affine group is a subgroup of the affine group, the "positive half" of the affine group. See Figure 8.

12 THE ORBITS

Consider how the positive affine group G affects Euclidean space R^n. The transformations $[a, c]$ carry a point $\mathbf{x}$ into the *orbit of* $\mathbf{x}$:

$$\begin{aligned} G\mathbf{x} &= \{[a, c]\mathbf{x}: \quad -\infty < a < \infty, \quad 0 < c < \infty\} \\ &= \{a\mathbf{1} + c\mathbf{x}: \quad -\infty < a < \infty, \quad 0 < c < \infty\}. \end{aligned}$$

The orbit is a half-plane, the half-plane passing through $\mathbf{x}$ bordered by the extended **1**-vector, but not including the extended **1**-vector. See Figure 9. For the special case of a point $\mathbf{x}$ on the line through the **1**-vector, the orbit is that line; this orbit with its special form can be excluded from R^n with no essential loss of generality in the sequel. Two orbits are either identical or disjoint: the orbits form a *partition* of the space R^n (see Problem 23).

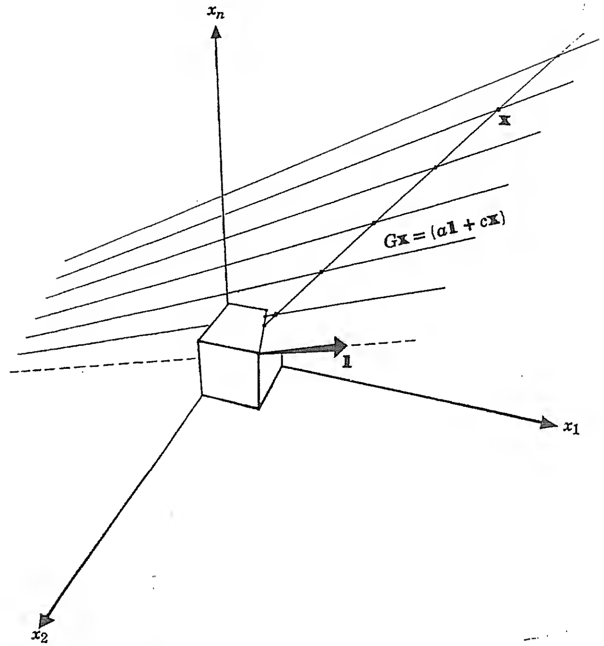


Figure 9 The orbit of $\mathbf{x}$ under the positive affine group.

Alternatively, the effect of the group can be examined by considering n numbered points $x_1, \ldots, x_n$ on the real line. A transformation $[a, c]$ carries these points into the n numbered points $\tilde{x}_1, \ldots, \tilde{x}_n$, where $\tilde{x}_i = a + cx_i$. The order of the points is unchanged, and the *relative* spacings between the points are unchanged. Only the *location* and *scaling* of the array of points are changed. See Figure 10.

Consider a simple variable to describe the *position* of a point $\mathbf{x}$ on its orbit (or the *location* and *scaling* of n numbered points on the real line). As an

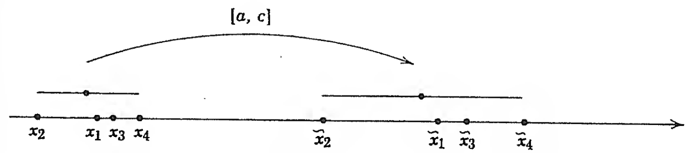


Figure 10 A transformation $[a, c]$ applied to four points.

exploratory choice consider $\bar{x}$ to measure location, s_x to measure scaling, and $[\bar{x}, s_x]$ to measure position;

$$s_x = \left[\frac{1}{n-1}\sum (x_i - \bar{x})^2\right]^{1/2}.$$

A transformation $[a, c]$ carries the point $\mathbf{x}$ into the point $\tilde{\mathbf{x}} = [a, c]\mathbf{x}$. The effect on the variables $\bar{x}$, s_x is

$$\bar{\tilde{x}} = a + c\bar{x},$$

$$s_{\tilde{x}} = cs_x,$$

which can be written

$$[\bar{\tilde{x}}, s_{\tilde{x}}] = [a, c][\bar{x}, s_x].$$

Thus, the transformation carries $\mathbf{x}$ into $[a, c]\mathbf{x}$ and correspondingly carries $[\bar{x}, s_x]$ into $[a, c][\bar{x}, s_x]$. See Figure 11.

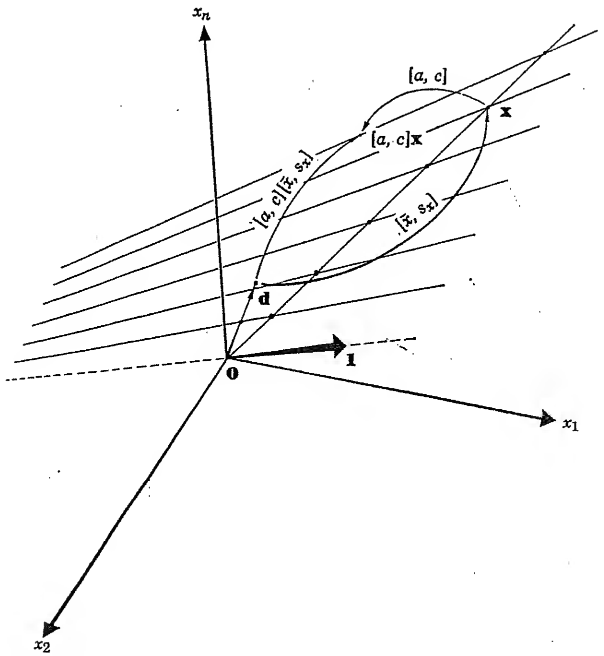


Figure 11 A transformation $[a, c]$ and its effect on the variable $[\bar{x}, s_x]$.

The inverse transformation $[\bar{x}, s_x]^{-1}$ can be used to reduce a general point $\mathbf{x}$ to a standard or *reference point* on its orbit:

$$[\bar{x}, s_x]^{-1}\mathbf{x} = \left(\frac{x_1 - \bar{x}}{s_x}, \ldots, \frac{x_n - \bar{x}}{s_x}\right)' = \mathbf{d}(\mathbf{x}).$$

The reference point has location $\bar{d} = 0$ and scale $s_d = 1$:

$$[\bar{d}, s_d] = [\bar{x}, s_x]^{-1}[\bar{x}, s_x] = [0, 1] = i.$$

The reference points $\mathbf{d}(\mathbf{x})$ *index* the orbits $G\mathbf{x}$.

The general point $\mathbf{x}$ can be reconstructed from its position and its orbit:

$$\mathbf{x} = [\bar{x}, s_x]\mathbf{d}(\mathbf{x}).$$

The variable $[\bar{x}, s_x]$ takes values in the group G. The variable $[\bar{x}, s_x]$, however, is used for *position* rather than for *transformation*; accordingly the range of $[\bar{x}, s_x]$ is designated G^* to distinguish the special use.

Now consider in general a variable to describe position on an orbit.

Definition 3. $[b(\mathbf{x}), s(\mathbf{x})]$ is a transformation variable if

$$[b([a, c]\mathbf{x}), s([a, c]\mathbf{x})] = [a, c][b(\mathbf{x}), s(\mathbf{x})]$$

for all $\mathbf{x}$, a, c; or equivalently if

$$b([a, c]\mathbf{x}) = a + cb(\mathbf{x}),$$
$$s([a, c]\mathbf{x}) = cs(\mathbf{x})$$

for all $\mathbf{x}$, a, c. $(c > 0)$.

As examples consider

$$\left[\frac{(\max x_i + \min x_i)}{2}, \max x_i - \min x_i\right],$$

$$[x_{(1)}, x_{(n)} - x_{(1)}],$$

where $x_{(i)}$ designates the ith smallest value of $x_1, \ldots, x_n$.

A transformation variable $[b(\mathbf{x}), s(\mathbf{x})]$ leads to a reference point on each orbit, the point at which the variable equals the identity:

$$\mathbf{d}(\mathbf{x}) = [b(\mathbf{x}), s(\mathbf{x})]^{-1}\mathbf{x} = \left(\frac{x_1 - b(\mathbf{x})}{s(\mathbf{x})}, \ldots, \frac{x_n - b(\mathbf{x})}{s(\mathbf{x})}\right)',$$

$$[b(\mathbf{d}), s(\mathbf{d})] = [b(\mathbf{x}), s(\mathbf{x})]^{-1}[b(\mathbf{x}), s(\mathbf{x})] = [0, 1].$$

The reference points $\mathbf{d}(\mathbf{x})$ index the orbits $G\mathbf{x}$.

The general point $\mathbf{x}$ can be reconstructed from position and orbit:

$$\mathbf{x} = [b(\mathbf{x}), s(\mathbf{x})]\mathbf{d}(\mathbf{x}).$$

Two transformation variables are simply related: *On any orbit they differ by right multiplication by a group element:*

$$[b_1(\mathbf{x}), s_1(\mathbf{x})] = [b_2(\mathbf{x}), s_2(\mathbf{x})][a, c].$$

The group element in general depends on the orbit; see Problem 24.

Now consider the measurement model

$$\prod f(e_i) \prod de_i,$$
$$\mathbf{x} = [\mu, \sigma]\mathbf{e}.$$

The points $\mathbf{x}$ and $\mathbf{e}$ are on the same orbit:

$$G\mathbf{x} = G\mathbf{e} \quad \text{or} \quad \mathbf{d}(\mathbf{x}) = \mathbf{d}(\mathbf{e}).$$

The positions of the points $\mathbf{x}$ and $\mathbf{e}$ differ by a transformation $[\mu, \sigma]$:

$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e].$$

The measurement model can then be reexpressed with composite structural equation:

$$\prod f(e_i) \prod de_i,$$
$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e], \qquad G\mathbf{x} = G\mathbf{e}.$$

Or, with transformation variable $[b(\mathbf{x}), s(\mathbf{x})]$, it can be reexpressed as

$$\prod f(e_i) \prod de_i,$$
$$[b(\mathbf{x}), s(\mathbf{x})] = [\mu, \sigma][b(\mathbf{e}), s(\mathbf{e})], \qquad G\mathbf{x} = G\mathbf{e}.$$

13 HOMOGENEITY

Consider a positive affine transformation $[a, c]$ and view the transformation as providing *new* coordinates for *given* points. The transformation rescales by the factor c and then relocates by the amount a:

$$\tilde{x} = a + cx.$$

The observation vector becomes

$$\tilde{\mathbf{x}} = [a, c]\mathbf{x}.$$

The scale characteristic σ becomes $\tilde{\sigma} = c\sigma$, and the response level μ becomes $\tilde{\mu} = a + c\mu$; accordingly, the system characteristic $[\mu, \sigma]$ becomes

$$[\tilde{\mu}, \tilde{\sigma}] = [a, c][\mu, \sigma].$$

The transformation does not touch the physical problem being examined; it affects only the numerical description of the quantities involved.

Consider the effect of the transformation on the model. The structural equation

$$\mathbf{x} = [\mu, \sigma]\mathbf{e}$$

can be multiplied on the left by $[a, c]$; it becomes

$$\tilde{\mathbf{x}} = [\tilde{\mu}, \tilde{\sigma}]\mathbf{e}.$$

Thus the model

$$\prod f(e_i) \prod de_i,$$
$$\mathbf{x} = [\mu, \sigma]\mathbf{e},$$

in the original coordinates becomes

$$\prod f(e_i) \prod de_i,$$
$$\tilde{\mathbf{x}} = [\tilde{\mu}, \tilde{\sigma}]\mathbf{e},$$

in terms of the new coordinates. The physical problem is untouched by the transformation $[a, c]$; the model conforms and has the same form after the transformation. The model is *homogeneous under the positive affine group.*

14 REDUCTION

Consider an application of the measurement model

$$\prod f(e_i) \prod de_i,$$
$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e], \qquad G\mathbf{x} = G\mathbf{e}.$$

The error distribution $\Pi f(e_i) \Pi de_i$ describes the internal error of the system as it affects the response; it describes the *random process* that generated the realized error $\mathbf{e}$ in the structural equation. The structural equation gives the relation between the known value $\mathbf{x}$ and the unknown values μ, σ, $\mathbf{e}$.

Now suppose the system is being examined in isolation, with no outside information concerning the unknowns; and consider the information in the structural equation concerning the unknowns μ, σ, $\mathbf{e}$. See Figure 12.

The orbit of $\mathbf{e}$ is known: $G\mathbf{e} = G\mathbf{x}$. And the information about the orbit is in the form of an *event* based on the variable $G\mathbf{e}$ for the random process $\mathbf{e}$.

Now consider the position of $\mathbf{e}$ on its orbit. The second part of the structural equation can be solved; it gives

$$[\bar{e}, s_e] = [\mu, \sigma]^{-1}[\bar{x}, s_x] = [A, C][\bar{x}, s_x].$$

This equation represents the position of $\mathbf{e}$ as an unknown transformation $[A, C]$ applied to $[\bar{x}, s_x]$. If the known position were different, $[\bar{\tilde{x}}, s_{\tilde{x}}] = [a, c][\bar{x}, s_x]$ for example, then the structural equation would give

$$[\bar{e}, s_e] = [\mu, \sigma]^{-1}[a, c][\bar{x}, s_x] = [A, C][\bar{x}, s_x],$$

where $[A, C] = [\mu, \sigma]^{-1}[a, c]$ is again an unknown transformation (homogeneity of the model). *Different* values for the position would provide the *same* description for $[\bar{e}, s_e]$. There is thus no information from the structural equation concerning the position of $\mathbf{e}$ on its orbit.

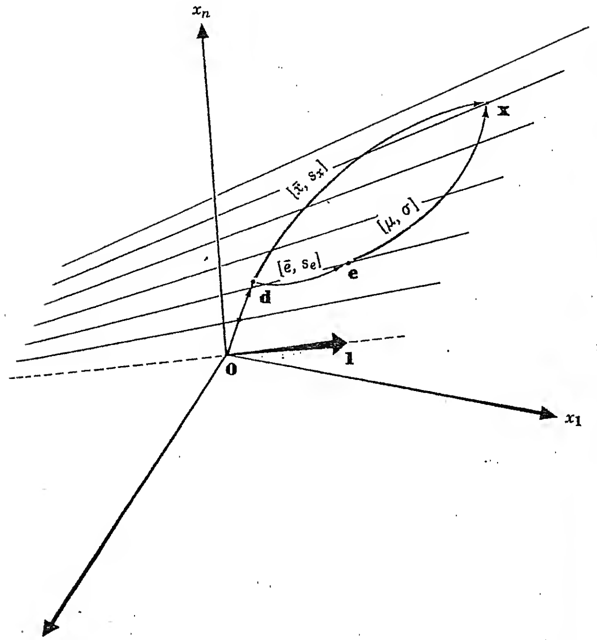


Figure 12 The known $[\bar{x}, s_x]$; the unknowns $[\bar{e}, s_e]$, $[\mu, \sigma]$.

IN SUMMARY. The error distribution describes the random process that generated the unknown $\mathbf{e}$ in the structural equation. The only other information concerning the unknown $\mathbf{e}$ has the form $G\mathbf{e} = G\mathbf{x}$, an event for the random process that generated $\mathbf{e}$. The conditions are fulfilled for making probability statements concerning unknown constants—exact probability statements can be made concerning the unknown error $\mathbf{e}$; they are based on the conditional distribution of the error variable $\mathbf{e}$ given the orbit $G\mathbf{e} = G\mathbf{x}$.

15 THE REDUCED MODEL

Consider the derivation of the conditional distribution of the error variable $\mathbf{e}$ given the value of the orbit $G\mathbf{e}$. The standard method used in Section 7 would proceed as follows: The two variables $\bar{e}, s_e$ describing position are supplemented by $n - 2$ variables describing the orbit; the Jacobian to the new variables is calculated; the joint density for the new variables is derived; the conditional density is obtained by normalizing over the variables $\bar{e}, s_e$. Consider instead a method based on the transformations that generate the orbits. This alternative method is easier here, even with the added explanation necessary for its introduction; for more complex problems it is simple and direct.

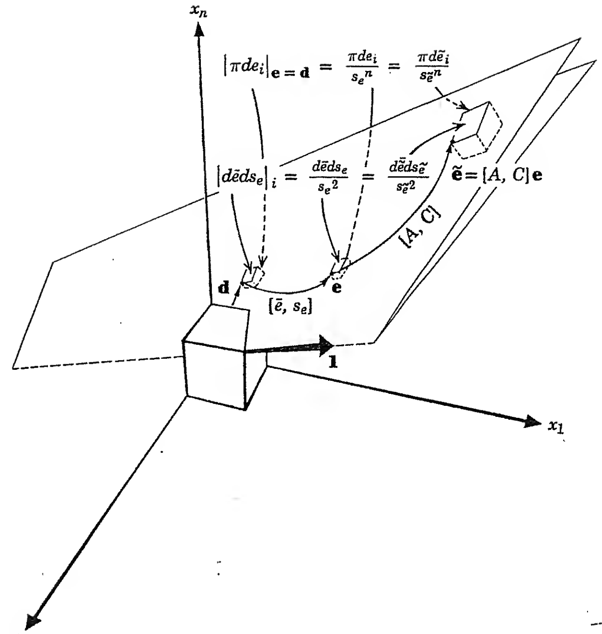


Figure 13 An element V at the reference point $\mathbf{d}$ and its images $[A, C]V$ under the group G.

Consider a neighborhood or *element* V at the reference point $\mathbf{d}(\mathbf{x}) = \mathbf{d}$. And consider the effect of transformations in G on this element. See Figure 13. The transformations $[A, C]$ carry this element point-for-point along orbits: *Position* is changed but *orbit* is not changed.

Consider first the effect of transformations in G as applied to the coordinates of R^n. The transformation

$$\tilde{\mathbf{e}} = [A, C]\mathbf{e}$$

is a diagonal transformation,

$$\begin{aligned} \tilde{e}_1 &= A + Ce_1, \\ &\cdot \\ &\cdot \\ &\cdot \\ \tilde{e}_n &= A \qquad + Ce_n, \end{aligned}$$

with Jacobian determinant

$$\left|\frac{\partial \tilde{e}}{\partial e}\right| = C^n;$$

the transformation $[A, C]$ applied to a point $\mathbf{e}$ changes Euclidean volume in R^n by the factor C^n. The special transformation $[\bar{e}, s_e]$ applied to the element V at the reference point $\mathbf{d}$ changes volume by the factor s_e^n. This factor can be used as a *compensating factor* to produce an adjusted differential

$$dm(\mathbf{e}) = \frac{\prod de_i}{s_e^n},$$

an adjusted differential that gives the *same* volume measure to all the images $[A, C]V$ of V:

$$\frac{\prod d\tilde{e}_i}{s_{\tilde{e}}^n} = \frac{C^n \prod de_i}{C^n s_e^n} = \frac{\prod de_i}{s_e^n}.$$

Note: A probability element such as $\Pi f(e_i) \Pi\, de_i$ is also an adjusted differential: the volume $\Pi\, de_i$ of an element at $\mathbf{e}$ is adjusted by the factor $\Pi f(e_i)$ to give the *probability* associated with that element. For the invariant differential $s_e^{-n} \Pi\, de_i$ the volume $\Pi\, de_i$ of an element at $\mathbf{e}$ is adjusted by the factor s_e^{-n} to give the volume of the corresponding element at the reference point $\mathbf{d}(\mathbf{e})$; the construction thus shows that this adjusted differential is the unique differential *that is invariant under the transformations and agrees with Euclidean volume at the reference points* $\mathbf{d}$ (where $s_d = 1$).

Consider the effect of transformations in G as applied to coordinates describing orbit and position on orbit. The transformations do not affect coordinates describing orbit: transformations carry points *along* orbits. Accordingly, *any* differential in terms of coordinates describing orbit is an *invariant* differential.

The transformations affect only position on orbit. The transformation

$$\bar{\tilde{e}} = A + C\bar{e}$$
$$s_{\tilde{e}} = Cs_e$$

is diagonal with Jacobian determinant C^2; the transformation $[A, C]$ changes Euclidean area (on G^*) by the factor C^2. The special transformation $[\bar{e}, s_e]$ applied to the element V at the identity $[0, 1]$ changes area by the factor s_e^2. (See Figure 14.) This factor can be used as a compensating factor to produce an adjusted differential

$$d\mu[\bar{e}, s_e] = \frac{d\bar{e}\, ds_e}{s_e^2}$$

that gives the *same* adjusted area to all the images $[A, C]V$ of V. The construction shows that this adjusted differential is the unique differential *that is*

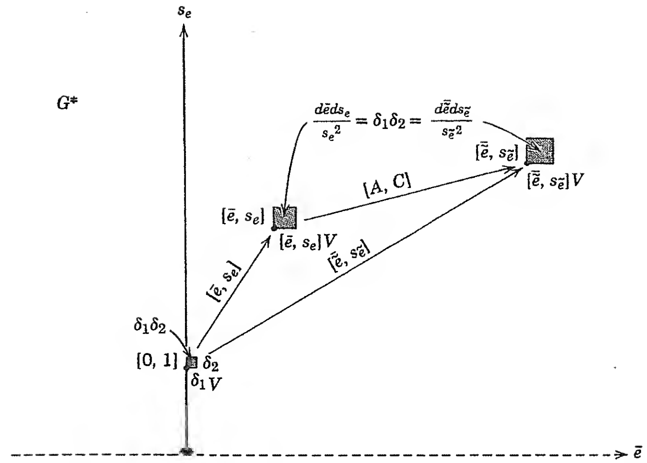


Figure 14 The volume element V and its images as viewed in terms of coordinates $[\bar{e}, s_e]$ in G^*.

invariant under the transformations and agrees with Euclidean area at the identity $[0, 1]$.

Now consider the two invariant differentials as they apply to the element V in R^n. Let $\delta(\mathbf{d})$ be their ratio at the reference point:

$$dm(\mathbf{e}) = \frac{\prod de_i}{s_e^n} = \delta(\mathbf{d})\frac{d\bar{e}\, ds_e}{s_e^2} = \delta(\mathbf{d})\, d\mu[\bar{e}, s_e].$$

The differentials, however, are unaffected by transformations $[A, C]$. The equality then holds generally, as the element V is transformed along orbits; the ratio $\delta(\mathbf{d})$ is a differential that measures V at right angles to the orbit. This provides the change of variables from a *volume element in the original coordinates to a volume element in position coordinates conditional on a neighborhood of the orbit* $\mathbf{d}(\mathbf{e}) = \mathbf{d}$:

$$\frac{\prod de_i}{s_e^n} = \delta(\mathbf{d})\frac{d\bar{e}\, ds_e}{s_e^2}.$$

The probability element for $\mathbf{e}$ on R^n is

$$\prod f(e_i) \prod de_i = \prod f(e_i) s_e^n \frac{\prod de_i}{s_e^n}.$$

In terms of the position variable $[\bar{e}, s_e]$ conditional on a neighborhood of the orbit $\mathbf{d}(\mathbf{e}) = \mathbf{d}$, the probability element becomes

$$\prod f(e_i) s_e^n\, \delta(\mathbf{d}) \frac{d\bar{e}\, ds_e}{s_e^2} = \prod f([\bar{e}, s_e] d_i) s_e^{n-2}\, \delta(\mathbf{d})\, d\bar{e}\, ds_e.$$

The conditional probability element for $[\bar{e}, s_e]$ given $\mathbf{d}(\mathbf{e}) = \mathbf{d}$ is obtained by normalization:

$$g(\bar{e}, s_e : \mathbf{d})\, d\bar{e}\, ds_e = k(\mathbf{d}) \prod f([\bar{e}, s_e] d_i) s_e^{n-2}\, d\bar{e}\, ds_e.$$

The constant $k(\mathbf{d})$ is the normalizing constant:

$$k^{-1}(\mathbf{d}) = \int_0^\infty \int_{-\infty}^\infty \prod f(\bar{e} + s_e d_i) s_e^{n-2}\, d\bar{e}\, ds_e.$$

The conditional distribution described in the preceding section has now been derived. The *measurement model* by its own information content produces the

Reduced Measurement Model

$$g(\bar{e}, s_e : \mathbf{d}(\mathbf{x}))\, d\bar{e}\, ds_e,$$

$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e].$$

The model has two parts: an *error probability distribution* $g(\bar{e}, s_e : \mathbf{d})\, d\bar{e}\, ds_e$ (with $[\bar{e}, s_e]$ as *variable* in the upper half-plane G^*) which provides probability statements for the unknown $[\bar{e}, s_e]$ in the structural equation; and a *structural equation* which gives the relation between the known $[\bar{x}, s_x]$ in G^* and the unknowns $[\mu, \sigma]$ in G and $[\bar{e}, s_e]$ in G^* (with $[\bar{e}, s_e]$ as a *constant*).

Any two position variables on an orbit are related by right multiplication by a group element; see Problem 24. The conditional probability element for a general position variable has then the form

$$g(b, s : \mathbf{d})\, db\, ds = k(\mathbf{d}) \prod f([b, s] d_i) s^{n-2}\, db\, ds,$$

where $\mathbf{d} = \mathbf{d}(\mathbf{x}) = [b(\mathbf{x}), s(\mathbf{x})]^{-1}\mathbf{x}$ is the reference point on $G\mathbf{x}$ and $k(\mathbf{d})$ is the normalizing constant.

16 EXAMPLES

As a first example consider the measurement model with standard normal error variable:

$$\prod f(e_i) \prod de_i = (2\pi)^{-n/2} \exp\{-\tfrac{1}{2} \sum e_i^2\} \prod de_i,$$

$$\mathbf{x} = [\mu, \sigma]\mathbf{e}.$$

The transformation variable $[\bar{e}, s_e]$ is convenient for the case of normal error. The conditional distribution of $[\bar{e}, s_e]$ given

$$\mathbf{d} = \left(\frac{e_1 - \bar{e}}{s_e}, \ldots, \frac{e_n - \bar{e}}{s_e}\right)'$$

is

$$g(\bar{e}, s_e : \mathbf{d})\, d\bar{e}\, ds_e = k(\mathbf{d})(2\pi)^{-n/2} \exp\{-\tfrac{1}{2} \sum (\bar{e} + s_e d_i)^2\} s_e^{n-2}\, d\bar{e}\, ds_e$$
$$= k' \exp\{-\tfrac{1}{2}[n\bar{e}^2 + (n-1)s_e^2]\} s_e^{n-2}\, d\bar{e}\, ds_e.$$

The simplification in the exponent uses $\Sigma d_i = 0$, $\Sigma d_i^2 = n - 1$ ($\bar{d} = 0$, $s_d = 1$).

$$g(\bar{e}, s_e : \mathbf{d})\, d\bar{e}\, ds_e = k'' \exp\left\{-\frac{n\bar{e}^2}{2}\right\} d\bar{e} \cdot \exp\left\{-\frac{(n-1)s_e^2}{2}\right\} (s_e^2)^{(n-1)/2-1}\, ds_e^2$$
$$= \left(\frac{n}{2\pi}\right)^{1/2} \exp\left\{-\frac{n\bar{e}^2}{2}\right\} d\bar{e}$$
$$\cdot \frac{1}{\Gamma((n-1)/2)} \left(\frac{(n-1)s_e^2}{2}\right)^{(n-1)/2-1} \exp\left\{-\frac{(n-1)s_e^2}{2}\right\} d\,\frac{(n-1)s_e^2}{2}.$$

The density factors so that the variables separate; the two factors are of normal and chi-square form; the usual normal and chi-square normalizing constants are introduced.

The conditional error distribution has the form: *$\bar{e}$ is normal with mean* 0 *and variance* $1/n$; $(n-1)s_e^2$ *is chi-square on* $n - 1$ *degrees of freedom*; *$\bar{e}$ and s_e are statistically independent.* It is of interest that the *conditional* distribution of $\bar{e}$, s_e *does not* depend on the value of $\mathbf{d}$; the *conditional* distribution of $\bar{e}$, s_e is thus the same as the *marginal* distribution of $\bar{e}$, s_e.

A chi-square density function on f degrees of freedom can be manipulated:

$$\frac{1}{\Gamma(f/2)} \exp\left\{-\frac{\chi^2}{2}\right\} \left(\frac{\chi^2}{2}\right)^{f/2-1} d\,\frac{\chi^2}{2} = \frac{A_f}{(2\pi)^{f/2}} \chi^{f-1} \exp\left\{-\frac{\chi^2}{2}\right\} d\chi,$$

where

$$A_f = \frac{2\pi^{f/2}}{\Gamma(f/2)}.$$

The conditional (or marginal) distribution can then be written

$$g(\bar{e}, s_e : \mathbf{d})\, d\bar{e}\, ds_e = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{n\bar{e}^2}{2}\right\} d\sqrt{n}\,\bar{e}$$
$$\cdot \frac{A_{n-1}}{(2\pi)^{(n-1)/2}} (\sqrt{n-1}\, s_e)^{n-2} \exp\left\{-\frac{(n-1)s_e^2}{2}\right\} d\sqrt{n-1}\, s_e.$$

The density, as a *marginal* density, is the original density

$$\frac{1}{(2\pi)^{n/2}} \exp\{-\tfrac{1}{2}\sum e_i^2\} = \frac{1}{(2\pi)^{n/2}} \exp\{-\tfrac{1}{2}n\bar{e}^2 - \tfrac{1}{2}(n-1)s_e^2\}$$

multiplied by the factor

$$A_{n-1}(\sqrt{n-1}\, s_e)^{n-2}.$$

The factor must give the result of integrating over the $(n-2)$-dimensional region corresponding to a *value* for the variable

$$(\sqrt{n}\,\bar{e}, \sqrt{n-1}\, s_e),$$

the original density being *constant* on this region. The region is *in* the $(n-1)$-dimensional linear subspace corresponding to a value for $\sqrt{n}\,\bar{e} = \Sigma\, e_i/\sqrt{n}$,

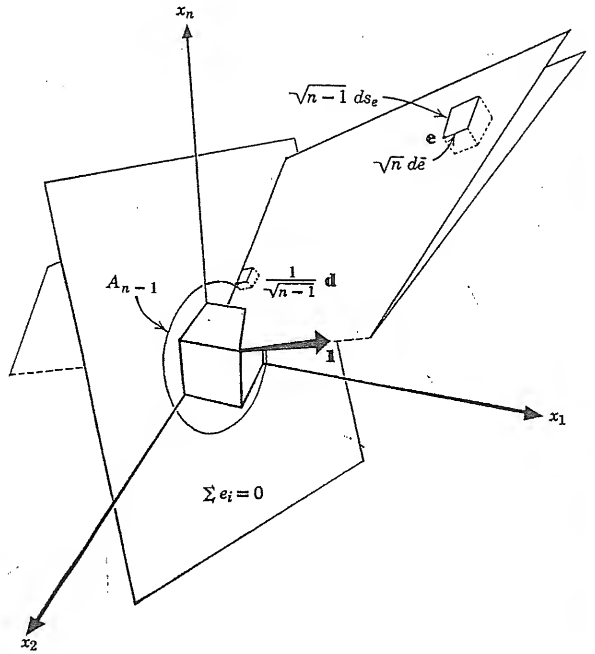


Figure 15 The invariant differential using coordinates in R^n and coordinates in the group. The area A_{n-1} of a unit sphere in the $(n-1)$-dimensional subspace $\sum e_i = 0$. The variable $\sqrt{n}\,\bar{e}$ measures distance in R^n parallel to the **1**-vector, and $\sqrt{n-1}\, s_e$ measures distance in R^n orthogonal to the **1**-vector.

that is, corresponding to a value of $e_1 + \cdots + e_n$. In this subspace it corresponds to a value for $(n-1)s_e^2 = \Sigma\,(e_i - \bar{e})^2$. The region is a *sphere* of radius $\sqrt{n-1}\, s_e$ in the $(n-1)$-dimensional linear subspace. See Figure 15. Thus the factor gives the *area of a sphere of radius* $\sqrt{n-1}\, s_e$ in $n-1$ dimensions, and A_{n-1} gives the *area of a unit sphere* in $n-1$ dimensions. This derivation of the area A_f of a unit sphere in f dimensions uses: *the invariant differentials* as derived from Jacobians (involving local properties); and the *one-dimensional integrations* that give the normalizing constants for the normal and gamma density functions. *No integration is needed in more than one dimension.*

The reduced model can now be expressed in the compact form:

$$\bar{e} = \frac{z}{\sqrt{n}}, \qquad s_e = \frac{\chi_{n-1}}{\sqrt{n-1}},$$
$$\bar{x} = \mu + \sigma\bar{e},$$
$$s_x = \sigma s_e.$$

The error distribution is described by means of a standard normal variable z and a chi-variable χ_{n-1} on $n-1$ degrees of freedom.

Suppose the measurements are

$$62.0, \quad 60.5, \quad 60.7, \quad 61.6.$$

The measurement model is

$$e_1 = z_1, \qquad e_2 = z_2, \qquad e_3 = z_3, \qquad e_4 = z_4,$$
$$62.0 = \mu + \sigma e_1,$$
$$60.5 = \mu + \sigma e_2,$$
$$60.7 = \mu + \sigma e_3,$$
$$61.6 = \mu + \sigma e_4.$$

The position values are $\bar{x} = 61.2$, $s_x = 0.72$; accordingly, the reduced model is

$$\bar{e} = \frac{z}{\sqrt{4}}, \qquad s_e = \frac{\chi_3}{\sqrt{3}},$$
$$61.2 = \mu + \sigma\bar{e},$$
$$0.72 = \sigma s_e.$$

The error probability distribution with $\bar{e}$, s_e as variables gives probability statements for the unknown error values $(\bar{e}, s_e)$ in the structural equation.

For example,

$$\Pr\{-1 \le \bar{e} \le 1\} = 95\tfrac{1}{2}\%,$$

$$\Pr\left\{\frac{\sqrt{0.216}}{\sqrt{3}} \le s_e \le \frac{\sqrt{9.35}}{\sqrt{3}}\right\} = 95\%.$$

As a second example consider the measurement model with error variable uniformly distributed on the interval (0, 1):

$$\prod f(e_i) \prod de_i,$$

$$\mathbf{x} = [\mu, \sigma]\mathbf{e},$$

where

$$\begin{aligned} f(e) &= 1 \qquad 0 \le e \le 1, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

The transformation variable

$$[L, R] = [\min e_i, \max e_i - \min e_i]$$

leads to a simple form for the conditional distribution, and its choice avoids a later change of reference point to gain simplicity. The conditional error distribution for $[L, R]$ given

$$\mathbf{d} = \left(\frac{e_1 - L}{R}, \ldots, \frac{e_n - L}{R}\right)'$$

is

$$\begin{aligned} g(L, R:\mathbf{d})\, dL\, dR &= k(\mathbf{d}) f(L + Rd_1) \cdots f(L + Rd_n) R^{n-2}\, dL\, dR \\ &= n(n-1)\varphi(L, R) R^{n-2}\, dL\, dR. \end{aligned}$$

The *indicator function* φ gives the range of nonzero density:

$$\begin{aligned} \varphi(L, R) &= 1 \qquad 0 \le L \le L + R \le 1, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

See Figure 16. The normalizing constant is obtained by integration:

$$\begin{aligned} k^{-1}(\mathbf{d}) &= \int_0^\infty \int_{-\infty}^\infty \varphi(L, R) R^{n-2}\, dL\, dR = \int_0^1 \int_0^{1-R} R^{n-2}\, dL\, dR \\ &= \int_0^1 (R^{n-2} - R^{n-1})\, dR \\ &= \frac{1}{n-1} - \frac{1}{n} = \frac{1}{n(n-1)}. \end{aligned}$$

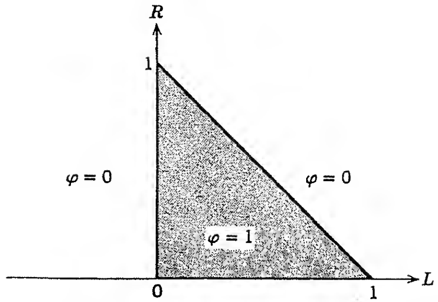


Figure 16 The region of positive density for the error position $[L, R]$.

The reduced model is

$$\begin{aligned} g(L, R:\mathbf{d})\, dL\, dR &= n(n-1)R^{n-2}\, dL\, dR \qquad 0 \le L \le L + R \le 1, \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

$$[L(\mathbf{x}), R(\mathbf{x})] = [\mu, \sigma][L, R].$$

17 TESTS OF SIGNIFICANCE

Consider the first example in the preceding section:

$$\bar{e} = \frac{z}{\sqrt{4}}, \qquad s_e = \frac{\chi_3}{\sqrt{3}},$$

$$61.2 = \mu + \sigma\bar{e},$$

$$0.72 = \sigma s_e.$$

Suppose that an outside source has prescribed the value $\mu_0 = 62.4$. The *hypothesis* $\mu = 62.4$ leads to information concerning the error:

$$61.2 = 62.4 + \sigma\bar{e},$$

$$0.72 = \qquad \sigma s_e,$$

$$\frac{\bar{e}}{s_e} = \frac{-1.2}{0.72}.$$

This value for $\bar{e}/s_e$ can be compared with the distribution of the error variable $\bar{e}/s_e$; or equivalently, the value

$$\frac{\sqrt{4}\,\bar{e}}{s_e} = \frac{2(-1.2)}{0.72} = -3.33$$

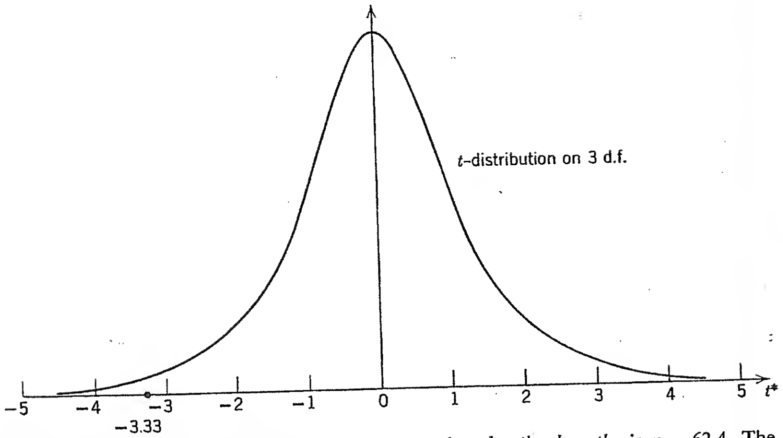


Figure 17 The error value $t^* = -3.33$ calculated under the *hypothesis* $\mu = 62.4$. The error probability distribution of t^*.

can be compared with the distribution of

$$\frac{\sqrt{4}\,\bar{e}}{s_e} = \frac{z}{\chi_3/\sqrt{3}} = t^*,$$

the t-distribution on three degrees of freedom. See Figure 17. The value -3.33 for t^* is just beyond the $2\frac{1}{2}\%$ point on the left-hand tail of the distribution; and it suggests moderately that the hypothesis is not true.

Now consider in general the measurement model

$$k(\mathbf{d}) \prod f([\bar{e}, s_e]d_i) s_e^{n-2}\, d\bar{e}\, ds_e,$$
$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e].$$

Suppose that an outside source has prescribed the value μ_0 for μ. The *hypothesis* $\mu = \mu_0$ gives

$$\bar{x} = \mu_0 + \sigma\bar{e},$$
$$s_x = \quad \sigma s_e,$$

and hence produces the value

$$t = \frac{\bar{e}}{s_e} = \frac{\sigma\bar{e}}{\sigma s_e} = \frac{\bar{x} - \mu_0}{s_x}$$

for the error characteristic t. This value for t can be compared with the distribution of t derived from the error probability distribution, $g(\bar{e}, s_e : \mathbf{d})\, d\bar{e}\, ds_e$: the joint probability element for t and s_e is

$$g(ts_e, s_e : \mathbf{d}) s_e\, dt\, ds_e;$$

the marginal element for t is then

$$\begin{aligned} g_L(t:\mathbf{d})\, dt &= \int_0^\infty g(ts_e, s_e : \mathbf{d}) s_e\, ds_e \cdot dt \\ &= k(\mathbf{d}) \cdot \int_0^\infty \prod f([ts_e, s_e]d_i) s_e^{n-1}\, ds_e \cdot dt \\ &= k(\mathbf{d}) \cdot \int_0^\infty \prod f(s_e(t + d_i)) s_e^{n-1}\, ds_e \cdot dt. \end{aligned}$$

The hypothesis can be assessed by comparing the calculated value of t with the distribution of values described by $g_L(t:\mathbf{d})\, dt$.

Now suppose, alternatively, that an outside source has indicated the value σ_0 for σ. The *hypothesis* $\sigma = \sigma_0$ leads to the value

$$s_e = \frac{s_x}{\sigma_0}$$

for the error variable s_e. This value for s_e can be compared with the distribution for s_e as obtained from the error probability distribution:

$$g_S(s_e : \mathbf{d})\, ds_e = k(\mathbf{d}) \cdot \int_{-\infty}^{\infty} \prod f([\bar{e}, s_e]d_i)\, d\bar{e} \cdot s_e^{n-2}\, ds_e;$$

and the hypothesis can be assessed accordingly.

18 GENERAL INFERENCE

Consider the measurement model and the information it contains concerning the unknown physical characteristic $[\mu, \sigma]$. For the numerical example in Section 16 it has the form

$$\bar{e} = \frac{z}{\sqrt{4}}, \qquad s_e = \frac{\chi_3}{\sqrt{3}},$$
$$61.2 = \mu + \sigma\bar{e},$$
$$0.72 = \quad \sigma s_e,$$

with error probability distribution that describes the unknown $[\bar{e}, s_e]$ and with structural equation that links the unknowns $[\bar{e}, s_e]$ and $[\mu, \sigma]$.

Each possible value for $[\bar{e}, s_e]$ corresponds to a possible value for $[\mu, \sigma]$:

$$[61.2, 0.72] = [\mu, \sigma][e, s_e].$$

A probability statement concerning $[\bar{e}, s_e]$ is *ipso facto* a probability statement concerning $[\mu, \sigma]$. The probability distribution that describes the unknown $[e, s_e]$ thus gives a distribution, the structural distribution, describing the unknown $[\mu, \sigma]$:

$$[\mu, \sigma] = [61.2, 0.72]\left[\frac{z}{\sqrt{4}}, \frac{\chi_3}{\sqrt{3}}\right]^{-1},$$

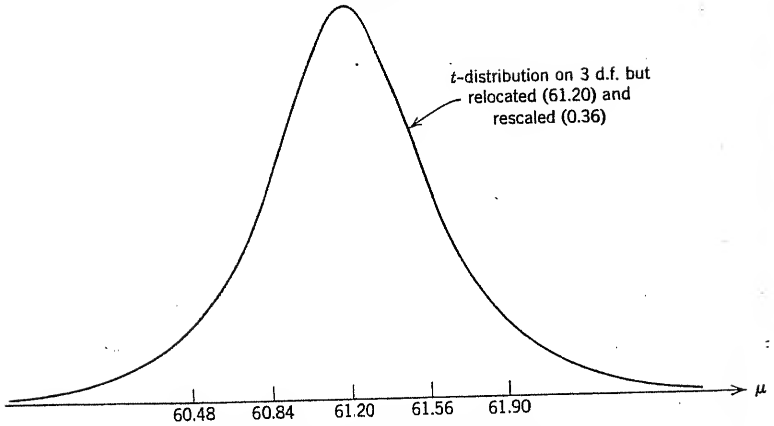


Figure 18 The structural distribution for μ.

or by coordinates

$$\mu = 61.2 - 0.72\frac{z/\sqrt{4}}{\chi_3/\sqrt{3}} = 61.2 - 0.36t_3^*,$$

$$\sigma = 0.72\frac{1}{\chi_3/\sqrt{3}}.$$

The distribution describing the unknown μ has the form of a t-distribution on three degrees of freedom, located at 61.2 and scaled by the factor $0.72/\sqrt{4} = 0.36$ (see Figure 18). The structural distribution for σ has the form of a χ^{-1} distribution with three degrees of freedom, scaled by the factor $0.72\sqrt{3}$.

In general for the measurement model with normal error, the reduced model is

$$\bar{e} = \frac{z}{\sqrt{n}}, \qquad s_e = \frac{\chi_{n-1}}{\sqrt{n-1}},$$

$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e].$$

The structural distribution is obtained by solving for $[\mu, \sigma]$:

$$[\mu, \sigma] = [\bar{x}, s_x][\bar{e}, s_e]^{-1};$$

$$\mu = \bar{x} - s_x\frac{\bar{e}}{s_e},$$

$$\sigma = \frac{s_x}{s_e};$$

and by using the error probability distribution to describe $[\bar{e}, s_e]$:

$$\mu = \bar{x} - \frac{s_x}{\sqrt{n}}\frac{z}{\chi_{n-1}/\sqrt{n-1}},$$

$$\sigma = \sqrt{n-1}\, s_x \cdot \chi_{n-1}^{-1}.$$

The *variables* μ, σ describing the structural distribution are statistically dependent, each variable involving χ_{n-1}.

Now consider the general model

$$k(\mathbf{d}) \prod f([\bar{e}, s_e]d_i)s_e^{n-2}\, d\bar{e}\, ds_e,$$

$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e].$$

The error probability distribution gives probability statements for the unknown $[\bar{e}, s_e]$. The structural equation links the values for $[\bar{e}, s_e]$ in one-to-one correspondence with the values for $[\mu, \sigma]$:

$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e] \quad \text{or} \quad [\bar{e}, s_e] = [\mu, \sigma]^{-1}[\bar{x}, s_x],$$

$$\bar{e} = \frac{\bar{x} - \mu}{\sigma}, \qquad s_e = \frac{s_x}{\sigma},$$

$$\left|\frac{\partial(\bar{e}, s_e)}{\partial(\mu, \sigma)}\right| = \begin{vmatrix} \dfrac{-1}{\sigma} & -\dfrac{(\bar{x} - \mu)}{\sigma^2} \\ 0 & -\dfrac{s_x}{\sigma^2} \end{vmatrix} = \frac{s_x}{\sigma^3}.$$

The distribution describing the unknown $[\bar{e}, s_e]$ gives the following *structural distribution* describing the unknown $[\mu, \sigma]$:

$$g^*(\mu, \sigma : \mathbf{x})\, d\mu\, d\sigma = k(\mathbf{d}) \prod f([\mu, \sigma]^{-1}x_i)\left(\frac{s_x}{\sigma}\right)^{n-2}\frac{s_x}{\sigma^3}\, d\mu\, d\sigma.$$

NOTES AND REFERENCES

The error variable of a stable system has been used as the basic ingredient to develop a statistical model and a method of inference. Other approaches to statistics examine the *exterior* of a system, and use the *classical model of statistics*: a possible input value θ denoting the physical quantity; an output value x denoting the observation; and a probability density $f(x:\theta)$ describing the frequency behavior of output values x for any input θ. The classical model effectively treats the system as a *black box*, a model that describes external behavior characteristics and ignores any internal operation or mechanisms. The other approaches need a variety of principles and

reduction techniques to obtain solutions to problems of inference. The approach here with a more comprehensive model obtains unique solutions without additional principles and techniques; the solutions are in terms of classical frequency-based probability.

The origins for the approach here lie somewhere between two diverse approaches to statistics, that of R. A. Fisher in England and that of the mathematical-statistical schools in North America. Both approaches use the classical model. The Fisher approach perhaps more frequently introduced methods stimulated by the peculiarities of the applications being studied.

Classical models can be derived from the models studied here: the location model $f(x - \theta)$ is obtained from the simple measurement model, and the location-scale model $\sigma^{-1}f((x - \mu)/\sigma)$ from the measurement model. Fisher examined these classical models in 1934 and proposed the use of a conditional distribution given a *configuration* of a sample. These *conditional models* are the classical-model analogs of the reduced models in this chapter.

Fisher (1930, 1935) also proposed distributions to describe the unknown value of a physical quantity. His proposal was in the framework of the classical model; his derivations violated generally accepted rules for handling probabilities and models.

Fisher R.A. (1930), Inverse probability, *Proc. Cambridge Phil. Soc.*, **26**, 528–535. Also as Paper 22, Fisher (1950).

Fisher R. A. (1934), Two new properties of mathematical likelihood, *Proc. Royal Soc. (London)*, **A144**, 285–307. Also as Paper 24, Fisher (1950).

Fisher R. A. (1935), The fiducial argument in statistical inference, *Ann. Eugenics*, **6**, 391–398. Also as Paper 25, Fisher (1950).

Fisher R. A. (1950), *Contributions to Mathematical Statistics*, Wiley, New York.

Fisher R. A. and Yates F. (1949), *Statistical Tables for Biological, Agricultural and Medical Research*, Hafner Publishing Co., New York.

Fraser D. A. S. (1961), The fiducial method and invariance, *Biometrika*, **48**, 261–280.

Fraser D. A. S. (1966), Structural probability and a generalization, *Biometrika*, **53**, 1–9.

Fraser D. A. S. (1968), A black box or a comprehensive model, *Technometrics*, **10**.

Pitman E. J. G. (1938), The estimation of the location and scale parameters of a continuous population of any given form, *Biometrika*, **30**, 391–421.

PROBLEMS

1. A strength measurement on batches of steel castings has error variation that is approximately normal with mean 0 and standard deviation 2.5 (units of 1000 psi). For a particular run of castings let θ be the general strength level. A random sample of 10 castings yielded

$$59.5,\quad 61.5,\quad 63.5,\quad 63.0,\quad 64.5,$$
$$61.5,\quad 60.0,\quad 65.0,\quad 59.5,\quad 57.0.$$

(i) Obtain the reduced model.

(ii) Make central 95% and 99% probability statements for the error position.

(iii) Derive the structural distribution for θ; make 95% and 99% probability statements for θ.

2. For the second example in Section 8 derive the structural distribution for θ; sketch the structural distribution.

3. Consider the simple measurement model with error variable uniformly distributed on the interval $(-0.5, 0.5)$.

(i) For the measurements $x_1 = 157.01$, $x_2 = 157.99$ derive the reduced model; derive the structural distribution for θ. *Note*. The conditional distribution can be obtained from the general formula or by geometrical argument from the uniform distribution of (e_1, e_2) over the square $(-0.5, 0.5) \times (-0.5, 0.5)$.

(ii) For the measurements 157.01, 157.99, 157.68, 157.92, 157.48 derive the reduced model; derive the structural distribution for θ. Do the additional three measurements add information concerning θ?

(iii) Test the hypothesis that $\theta = 157.60$.

4. Consider the simple measurement model with normal error in Section 8. Use the location variable $r(x) = \min x_i$ with $d_i = x_i - \min x_i$, and show that the conditional error distribution for $\min e_i$ is normal with mean $-\bar{d}$ and variance σ_0^2/n. Check to see if this is equivalent to the conditional error distribution in Section 8.

5. Consider the simple measurement model with component error density

$$f(e) = \exp\{-e\}, \qquad e > 0,$$
$$= 0, \qquad e \le 0.$$

(i) Derive the reduced model using $r(x) = \min x_i$ as location variable; determine the normalizing constant.

(ii) Derive the structural distribution for θ.

***6.** Consider the simple measurement model with double exponential component error:

$$f(e) = \tfrac{1}{2} \exp\{-|e|\}.$$

NOTATION. Let $x_{(i)}$ be the ith smallest of $(x_1, \ldots, x_n)$; then $x_{(1)} = \min x_i$, $x_{(n)} = \max x_i$; each $x_{(i)}$ is a location variable.

(i) Derive the reduced model using $x_{(1)}$ as location variable; determine the normalizing constant (integration can be performed interval by interval on the real line).

(ii) Derive the structural distribution for θ.

(iii) For the measurements 5.8, 6.5, 6.8 sketch the reduced error distribution; sketch the structural distribution.

7. Consider the card-dealing example in Section 5. Find Pr {2 spades} with the additional information that the second participant observed both cards and would not have differentially reported the special case of two spades.

8. Show that the positive affine transformations form a group:

$$G = \{[a, c]: \quad -\infty < a < \infty, \quad 0 < c < \infty\},$$

the *positive affine group* or *location-scale group*.

9. For the numerical example at the beginning of Section 17 test the hypothesis that $\sigma = 0.3$.

10. A method of measuring temperature remotely has an error variable that is approximately standard normal. For a particular sequence of seven determinations (°C),

$$683,\quad 688,\quad 683,\quad 687,\quad 692,\quad 687,\quad 682,$$

let μ be the temperature level and σ be the error scaling.

(i) Obtain the reduced model.
(ii) Make central 90% probability statements for the error characteristics $t = \bar{e}/s_e$, s_e.
(iii) Obtain and sketch the structural distribution for μ and the structural distribution for σ.
(iv) Make central 90% probability statements for μ and for σ.

11. Test the hypothesis: $\mu = 680°\text{C}$ in the preceding example.

12. Consider the measurement model with component error distribution

$$f(e) = \exp\{-e\}, \qquad e > 0,$$
$$= 0, \qquad e \le 0.$$

(i) Derive the reduced model using $[b, s] = [x_{(1)}, \sum_2^n x_{(i)}/(n-1) - x_{(1)}]$ as transformation variable.
(ii) Derive the distribution of the error characteristic $t = b/s$. Sketch the structural distribution of μ.

13. Consider the measurement model with error uniformly distributed on the interval $(0, 1)$.

(i) Derive the reduced model using $[b, s] = [x_{(1)}, x_{(n)} - x_{(1)}]$ as transformation variable.
(ii) Derive the distribution of the error characteristic $t = b/s$.
(iii) Derive the structural distribution for $[\mu, \sigma]$.

***14.** Consider the measurement model with double exponential component error

$$f(e) = \tfrac{1}{2}\exp\{-|e|\}.$$

(i) Derive the reduced model using $[b, s] = [x_{(1)}, x_{(2)} - x_{(1)}]$ as position variable.
(ii) For the measurements 5.8, 6.5, 6.8 sketch several sections of the conditional distribution: for example, the section $s = s_0$ and the section $b = t_0 s$.
(iii) Derive the distribution of the error characteristic $t = b/s$.
(iv) Sketch the structural distribution for the response level μ.

***15.** The general Weibull distribution is

$$\frac{\beta}{\alpha^\beta}(t-\gamma)^{\beta-1}\exp\left\{-\left(\frac{t-\gamma}{\alpha}\right)^\beta\right\}dt, \qquad t > \gamma,$$

with $0 < \alpha < \infty$, $0 < \beta < \infty$, $-\infty < \gamma < \infty$. Consider the measurement model with a Weibull component error distribution ($\beta > 0$ given):

$$f(e)\,de = \beta e^{\beta-1}\exp\{-e^\beta\}\,de, \qquad e > 0. \qquad \text{(J. Whitney.)}$$

(i) Derive the reduced model.
(ii) Derive the structural distribution for $[\mu, \sigma]$.

***16.** The general Weibull distribution can be specialized:

$$\frac{\beta}{\alpha^\beta}t^{\beta-1}\exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}dt = \frac{\beta}{\alpha^\beta}\exp\{\beta\ln t - \exp\{\beta(\ln t - \ln\alpha)\}\}\,d\ln t, \qquad t > 0.$$

Consider the measurement model with a Weibull component error distribution

$$f(e)\,de = \exp\{e - \exp\{+e\}\}\,de. \qquad \text{(J. Whitney.)}$$

(i) Derive the reduced model.
(ii) Derive the structural distribution for $[\mu, \sigma]$.
(iii) Derive the structural distribution for σ.

***17.** Consider a first measurement model with error distribution $f_1(e)$, observations $(x_1, \ldots, x_m)$, and quantities $[\mu_1, \sigma_1]$. And consider a second measurement model with error distribution $f_2(e)$, observations $(y_1, \ldots, y_n)$, and quantities $[\mu_2, \sigma_2]$.

(i) Derive an integral expression for the structural distribution of $\mu_1 - \mu_2$.
(ii) Now suppose that $f_1(e), f_2(e)$ are standard normal. Show that the structural distribution of $\mu_1 - \mu_2$ can be represented in the form

$$\mu_1 - \mu_2 = \bar{x} - \bar{y} + r(t_1\cos\theta + t_2\sin\theta),$$

where

$$r^2 = \frac{s_x^2}{m} + \frac{s_y^2}{n},$$

$$\tan\theta = \frac{s_y/\sqrt{n}}{s_x/\sqrt{m}},$$

and t_1 and t_2 are t-variables on $m - 1$ and $n - 1$ degrees of freedom, respectively. Tables for the distribution of a linear combination $t_1\cos\theta + t_2\sin\theta$ of t-variables have been prepared by Sukhatme and are tabulated in Fisher and Yates (1949). (Fisher, 1935.)

18. Show that the rescalings

$$G = \{[0, c]:\ 0 < c < \infty\}$$

form a group, the *scale group*.

19. Let e be an error variable with distribution $f(e)\,de$ on the real line R^1; let θ be a quantity taking positive values; and let $x_1, \ldots, x_n$ be measurements with a multiplicative error. The *multiplicative measurement model* is

$$\prod f(e_i)\prod de_i,$$
$$\mathbf{x} = [0, \theta]\mathbf{e} = \theta\mathbf{e},$$

where $[0, \theta]$ is an element of the scale group.

(i) Determine the orbit of $\mathbf{x}$ under the scale group (delete the origin $\mathbf{0}$).
(ii) Define a *scale variable* $s(\mathbf{x})$. Show that $(\sum x_i^2)^{1/2}$ is a scale variable.
(iii) Show that the orbits can be indexed by $\mathbf{d}(\mathbf{x}) = (x_1/s(\mathbf{x}), \ldots, x_n/s(\mathbf{x}))'$.
(iv) Derive the *reduced model*

$$k(\mathbf{d})s^n\prod f(sd_i)\cdot\frac{ds}{s},$$
$$s(\mathbf{x}) = \theta s,$$

where $s = s(\mathbf{e})$ designates error position.

(v) Derive the structural distribution for θ:

$$k(\mathbf{d})\left(\frac{s(\mathbf{x})}{\theta}\right)^n \prod f\left(\frac{x_i}{\theta}\right)\cdot\frac{d\theta}{\theta}.$$

(vi) Suppose that $f(e) = 0$ for $-\infty < e \le 0$: Show that a logarithmic transformation applied to the variables in this problem transforms the multiplicative measurement model into the simple measurement model.

20. Consider the multiplicative measurement model with standard normal error. Show that the structural distribution for θ is

$$\frac{A_n}{(2\pi)^{n/2}}\left(\frac{s(\mathbf{x})}{\theta}\right)^n \exp -\left\{\frac{1}{2\theta^2}s^2(\mathbf{x})\right\}\frac{d\theta}{\theta};$$

use $(\sum x_i^2)^{1/2}$ as scale variable.

21. Consider the multiplicative measurement model with error distribution uniform on the interval (0, 1).

(i) Derive the reduced model using $\max x_i = s(\mathbf{x})$ as scale variable.
(ii) Derive the structural distribution for θ.

22. A radioactive source emitting particles at the average rate of one per unit time interval gives the distribution

$$\begin{aligned} f(e)\,de &= \exp\{-e\}\,de, & e &> 0,\\ &= 0, & e &\le 0, \end{aligned}$$

for the time interval e between successive emissions. Let θ be the corresponding time interval for a source under investigation, and let $x_1, \ldots, x_n$ be n independent measurements of time interval between successive emissions. This is the multiplicative measurement model

$$\prod f(e_i) \prod de_i,$$
$$\mathbf{x} = [0, \theta]\mathbf{e}.$$

(i) Describe the orbits and reference points; use $\bar{x}$ as scale variable.
(ii) Derive the conditional error distribution $g(\tilde{e}:\mathbf{d})\,d\tilde{e}$.
(iii) Derive the structural distribution for θ.

23. For the positive affine group show that two orbits are either identical or disjoint.

24. (i) Show that the following variables are transformation variables for the positive affine group:

$$[b, s] = [x_1, |x_2 - x_1|],$$

$$[M, R] = \left[\frac{(\max x_i + \min x_i)}{2}, \max x_i - \min x_i\right];$$

for the variable $[b, s]$ the exceptional set must be increased from the extended 1-vector to the $(n-1)$-dimensional subspace described by $x_2 - x_1 = 0$.

(ii) For the two transformation variables $[\bar{x}, s_x]$ and $[M, R]$ determine the connecting transformation $[a, c]$,

$$[M, R] = [\bar{x}, s_x][a, c],$$

and show that $[a, c]$ is constant-valued on any orbit.

(iii) Let $[b_1(\mathbf{x}), s_1(\mathbf{x})]$ and $[b_2(\mathbf{x}), s_2(\mathbf{x})]$ be two transformation variables and $\mathbf{d}_1(\mathbf{x})$ and $\mathbf{d}_2(\mathbf{x})$ the corresponding reference points. Show that

$$[b_1(\mathbf{x}), s_1(\mathbf{x})] = [b_2(\mathbf{x}), s_2(\mathbf{x})][a, c],$$
$$[a, c]\mathbf{d}_1(\mathbf{x}) = \mathbf{d}_2(\mathbf{x}),$$

where $[a, c]$ depends only on the orbit, is constant-valued on any orbit. Draw a diagram of an orbit as a half-plane and indicate the significance of the equations.

*25. Consider the measurement model with position variable $[b, s] = [x_1, |x_2 - x_1|]$ and reference point

$$\mathbf{d}^* = \left(\frac{x_1 - b}{s}, \ldots, \frac{x_n - b}{s}\right)'.$$

(i) Use the method of Section 7 to derive the conditional error distribution.
(ii) Find the form of the conditional error distribution for the normal example at the beginning of Section 16; simplify by using the alternate coordinates $[\bar{e}, s_e]$ on the orbit (see Problem 24).

26. (i) The positive affine transformations $[a, c]$ can be reexpressed in terms of matrices. Show that the set

$$G = \left\{\begin{bmatrix} 1 & 0 \\ a & c \end{bmatrix} : \begin{array}{l} -\infty < a < \infty \\ 0 < c < \infty \end{array}\right\}$$

with matrix multiplication as the operation forms a group having effectively the same multiplication rule as the positive affine group.

(ii) Check that the measurement model can be reexpressed in terms of matrices and matrix multiplication:

$$E = \begin{bmatrix} 1 & \cdots & 1 \\ e_1 & \cdots & e_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix},$$
$$\theta = \begin{bmatrix} 1 & 0 \\ \mu & \sigma \end{bmatrix},$$
$$f(E)\,dE = \prod_1^n f(e_i) \prod de_i,$$
$$X = \theta E.$$

27. The positive affine transformations $[a, c]$ form a group

$$G = \left\{[a, c]: \begin{array}{l} -\infty < a < \infty \\ 0 < c < \infty \end{array}\right\}$$

with the multiplication rule

$$[a, c][a^*, c^*] = [a + ca^*, cc^*].$$

Let A be a $p \times r$ matrix of real numbers, and C be a $p \times p$ matrix of real numbers with positive determinant. Show that the elements $[A, C]$ form a group (the *generalized positive affine group* or the *regression positive-linear group*)

$$G = \{[A, C]\}:$$

the multiplication rule is

$$[A, C][A^*, C^*] = [A + CA^*, CC^*],$$

and the identity is $[0, I]$ where I is the $p \times p$ identity matrix.

28. The generalized affine transformations $[A, C]$ can be reexpressed in terms of $(p + r) \times (p + r)$ matrices. Show that the set

$$G = \left\{ \left(\underbrace{\begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix}}_{A} \left| \underbrace{\begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{matrix}}_{C} \right. \right) : |C| > 0 \right\},$$

with matrix multiplication as the operation forms a group having effectively the same multiplication rule as the generalized positive affine group.

29. A set $\{A_\alpha\}$ of subsets A_α of a space $\mathfrak{X}$ is a partition if (a) any two subsets $A_\alpha, A_\beta (\alpha \neq \beta)$ are disjoint: $A_\alpha \cap A_\beta = \varnothing$, where $\varnothing$ is the empty set; (b) *the union of all the sets* A_α *is the space* $\mathfrak{X}$: $\bigcup_\alpha A_\alpha = \mathfrak{X}$.

(i) Show that the orbits in Section 3 form a partition of R^n.
(ii) Show that the orbits in Section 12 form a partition of R^n (extended 1-vector deleted).

30. Show that a set G of one-to-one transformations of a set $\mathfrak{X}$ onto itself is a group if and only if it is *closed* under the formation of products and inverses: if g_1, g_2 are in G, then $g_1 g_2$ is in G; if g is in G, then g^{-1} is in G.

CHAPTER TWO

The Structural Model

In the development of the measurement model the internal error of a system was recognized as the primary entity. The error variable was introduced to describe the response effect of the internal error; it is the basic ingredient of the model.

The measurement model, however, corresponds to a rather special kind of system, a system with all controllable variables held constant and with a single real-valued response. This chapter introduces a general model, the *structural model*. The structural model corresponds to a general system with internal error, and it has an error variable to describe the response effect of the internal error.

The development of the structural model follows very closely the pattern in Chapter One. The two measurement models were analyzed there in a form that would best illustrate the general methods and concepts of this chapter. Some of these methods and concepts are trivial for the simple measurement model; all are nontrivial for the measurement model. The structural model is developed without further illustrations; some simple extensions of the measurement model are introduced in the Problems.

1 THE MODEL

Consider a system operating under stable conditions. Suppose that experience with such a system using appropriate measurement scales has led to the identification of a *response component of the internal error*. Let this be described by an *error variable* E having a fixed distribution on the space $\mathfrak{X}$ of the response.

Suppose the general characteristics of the system are given by a quantity θ, a transformation belonging to a group G of transformations of $\mathfrak{X}$ onto $\mathfrak{X}$. To avoid triviality the group G is assumed to be *unitary* on $\mathfrak{X}$:

Definition 1. A group G of transformations on $\mathfrak{X}$ is unitary if $g_1 x = g_2 x$ for any x implies $g_1 = g_2$.

A group is unitary if there is at most one transformation carrying any point into any other point. The quantity θ applied to a value E of the error variable gives a value X for the response: $X = \theta E$.

This description of the system produces the

Structural Model

$$E,$$
$$X = \theta E.$$

The structural model has two parts: an *error variable* E with a known distribution on the space $\mathfrak{X}$; and a *structural equation* $X = \theta E$ in which a *realized* value E from the error variable gives the relation between the known response X in $\mathfrak{X}$ and the unknown quantity θ in the group G of transformations on $\mathfrak{X}$.

Consider how the group G affects the space $\mathfrak{X}$. The transformations g in G carry a point X into the *orbit of* X:

$$GX = \{gX: \ g \in G\}.$$

(see Figure 1.)

Suppose two points are related by a transformation:

$$X_1 = hX_2, \qquad X_2 = h^{-1}X_1.$$

Then any point generated from one can be generated from the other:

$$gX_1 = (gh)X_2, \qquad \tilde{g}X_2 = (\tilde{g}h^{-1})X_1,$$

and the orbit of X_1 is the same as the orbit of X_2. It follows then that any two orbits are either identical or disjoint and that the orbits *partition* the space $\mathfrak{X}$ (Definition in Problem 29, Chapter One).

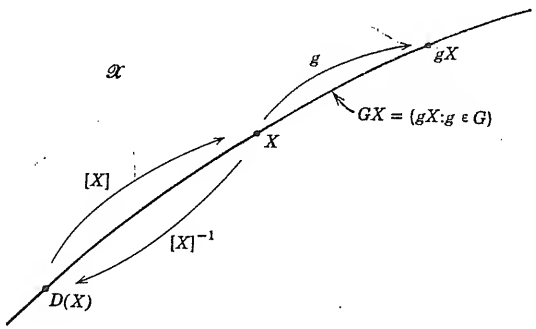


Figure 1 Orbit, reference point, and transformation variable.

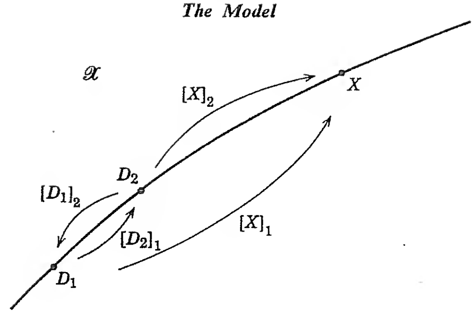


Figure 2 Comparing two transformation variables and the corresponding reference points.

Consider a variable to describe position on an orbit:

Definition 2. A function $[X]$ defined on $\mathfrak{X}$ and taking values in G is a transformation variable if

$$[gX] = g[X]$$

for all g in G and X in $\mathfrak{X}$.

A transformation variable $[X]$ leads to a *reference point* $D(X)$ on each orbit (the point at which the variable equals the identity):

$$D(X) = [X]^{-1}X,$$
$$[D(X)] = [X]^{-1}[X] = i.$$

A transformation variable can always be defined by choosing a reference point $D(X)$ on each orbit GX and letting $[X]$ be the unique element in G that carries the reference point into X:

$$X = [X]D(X).$$

The reference points *index* the orbits, and the transformation variable gives *position* on an orbit.

Two transformation variables are simply related one to the other. Along any orbit they differ by right multiplication by a group element:

$$X = [X]_1\, D_1(X), \qquad X = [X]_2\, D_2(X),$$
$$[X]_2 = [X]_1[D_1(X)]_2,$$
$$[X]_1 = [X]_2[D_2(X)]_1.$$

(See Figure 2.)

Consider the structural model again:

$$E,$$
$$X = \theta E,$$

and let $[X]$ be a transformation variable. The points X and E are on the same orbit:

$$GX = GE \quad \text{or} \quad D(X) = D(E).$$

The positions of the points X and E differ by a transformation θ:

$$[X] = \theta[E].$$

The structural model can then be rewritten with composite structural equation:

$$E,$$
$$[X] = \theta[E], \qquad GX = GE.$$

The quantity θ is an element of the group G. The positions $[E]$ and $[X]$ are also elements of the group but designated G^* to distinguish the use for *position* as opposed to *transformation*.

The structural model can be written alternatively

$$E,$$
$$[X] = \theta[E], \qquad D(X) = D(E).$$

Consider a transformation g and view it as providing new coordinates for given entities:

$$\tilde{X} = gX, \qquad \tilde{\theta} = g\theta.$$

The structural model in the original coordinates is

$$E,$$
$$X = \theta E.$$

Multiplying the structural equation by g gives $\tilde{X} = \tilde{\theta}E$. The structural model then becomes

$$E,$$
$$\tilde{X} = \tilde{\theta}E.$$

The model thus has the same form in the new coordinates $\tilde{X}$, $\tilde{\theta}$ as in the original coordinates X, θ: the model is *homogeneous under the group G*.

2 THE REDUCED MODEL

Consider an application of the structural model

$$E,$$
$$[X] = \theta[E], \qquad GX = GE,$$

to a system under stable conditions. The error variable E describes the response component of the internal error; it describes the essentials of the

random process that generated the realized E in the structural equation. The structural equation gives the relation between the known X and the unknowns θ and E.

Now suppose the system is being examined in isolation with no outside information concerning the unknowns; and consider the information in the structural equation concerning the unknowns θ and E.

The orbit of E is known: $GE = GX$. And the information about the orbit is in the form of an event based on the variable GE for the random process E.

The position of E on its orbit, however, is not known:

$$[E] = \theta^{-1}[X] = g[X];$$

the structural equation represents $[E]$ as an *unknown transformation* g applied to the known $[X]$. If the known position were different, $[\tilde{X}] = h[X]$ for example, then the description of $[E]$ would be

$$[E] = \theta^{-1}h[X] = g[X],$$

where g is again an *unknown transformation* in G. Different values for position would give the same description of $[E]$. There is thus no information in the structural equation concerning the position of E on its orbit.

The error distribution describes the random process that generated the unknown E in the structural equation. The only other information concerning the unknown E has the form $GE = GX$, an event for the random process that generated E. The conditions are fulfilled for making probability statements concerning unknown constants—exact probability statements can be made concerning the unknown error E; they are based on the conditional distribution of the error variable E, given the orbit $GE = GX$. The structural model by its own information content produces the

Reduced Structural Model

$$[E]: GE = GX,$$
$$[X] = \theta[E].$$

The reduced model has two parts: an *error probability distribution* (for the *variable* $[E]$ given $GE = GX$) which provides probability statements for the unknown position $[E]$ in the structural equation; and a *structural equation* which gives the relation between the known $[X]$ in G^* and the unknowns θ in G and $[E]$ in G^* (with $[E]$ as a *constant*). (See Figure 3.)

3 INVARIANT DIFFERENTIALS

In Chapter One the conditional error distribution for the measurement model was derived by means of invariant differentials. An element about an initial reference point was transformed along orbits; its volume was measured

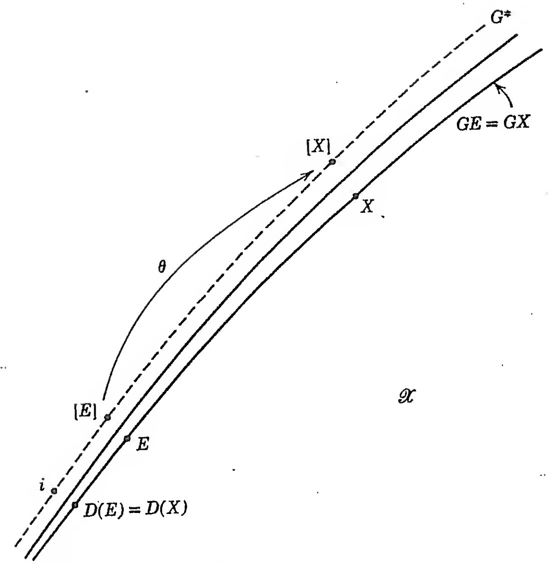


Figure 3 The orbit of X and E; the known $[X]$ in G^*, the unknowns θ in G and $[E]$ in G^*.

invariantly in terms of original coordinates and invariantly in terms of orbit and position coordinates; the equivalence of the two invariant measures gave the change of differential from original coordinates to position coordinates given orbit; the probability element (in terms of position coordinates given orbit) was normalized over the range of the position coordinates. This procedure is now used for the structural model.

Consider a unitary group G of one-to-one transformations of $\mathfrak{X}$ onto $\mathfrak{X}$ and suppose the following assumption holds:

Assumption 3.† $\mathfrak{X}$ is an open set in Euclidean space R^N; G is an open set in R^L; the transformations

$$\tilde{g} = hg, \qquad \tilde{X} = hgX$$

are continuously differentiable with respect to g, h, and X; and $[X]$ is a continuous transformation variable on $\mathfrak{X}$.

Consider a neighborhood or *element* V at a reference point $D(X) = [X]^{-1}X$; and consider the effect of transformations in G on this element (see

† The methods and results remain valid if $\mathfrak{X}$ and G are Euclidean manifolds; they also remain valid if $\mathfrak{X}$ and G are topological spaces provided derivatives are replaced by appropriate Nikodym derivatives relative to a given measure on $\mathfrak{X}$.

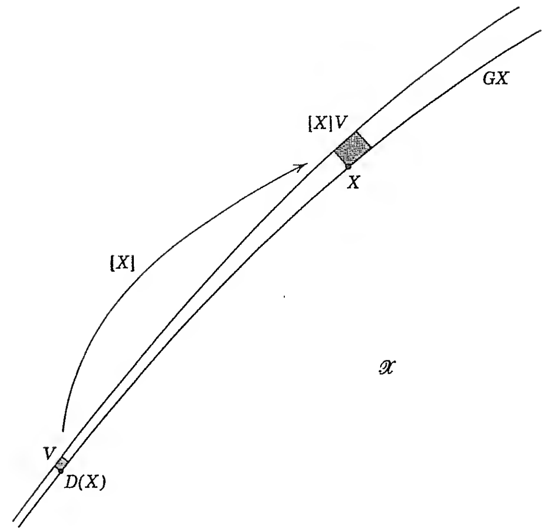


Figure 4 Transformations in G applied to an element V at the reference point $D(X)$.

Figure 4). A transformation h applied to a point X changes Euclidean volume (in R^N),

$$dhX = \left|\frac{\partial hX}{\partial X}\right| dX,$$

by the positive-Jacobian factor

$$J_N(h:X) = \left|\frac{\partial hX}{\partial X}\right|.$$

A composite transformation h_2h_1 applied to X changes volume by the factor

$$\begin{aligned} J_N(h_2h_1:X) &= \left|\frac{\partial h_2h_1X}{\partial X}\right| = \left|\frac{\partial h_2h_1X}{\partial h_1X}\right| \cdot \left|\frac{\partial h_1X}{\partial X}\right| \\ &= J_N(h_2:h_1X)J_N(h_1:X). \end{aligned}$$

The particular transformation $[X]$ applied to the reference point $D(X)$ changes volume by the factor

$$J_N(X) = J_N([X]:D),$$

which can be used as a *compensating factor* to produce an *invariant differential*

$$dm(X) = \frac{dX}{J_N(X)},$$

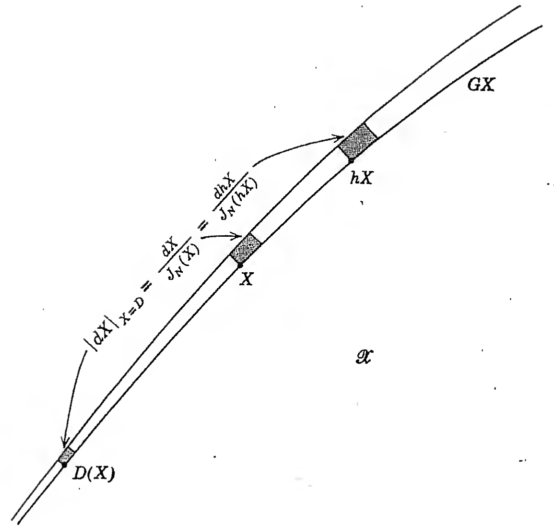


Figure 5 The invariant differential that coincides with dX at the reference point D.

a measure of volume that is constant under any transformation h in G:

$$dm(hX) = \frac{dhX}{J_N(hX)} = \frac{J_N(h:X)\,dX}{J_N(h:X)J_N(X)} = \frac{dX}{J_N(X)} = dm(X).$$

The construction shows that it is the *unique* invariant differential that coincides with Euclidean volume at the reference point $D(X)$. (See Figure 5.)

As an example consider the positive affine group with transformation variable $[\bar{x}, s_x]$ (Section 12, Chapter One):

$$d[A, C]\mathbf{x} = C^n\, d\mathbf{x},$$
$$J_n([A, C]: \mathbf{x}) = C^n,$$
$$J_n(\mathbf{x}) = s_x^n,$$
$$dm(\mathbf{x}) = \frac{d\mathbf{x}}{s_x^n}.$$

Consider now the effect of transformations in G as applied to coordinates describing orbit and position on orbit. The transformations do *not* affect coordinates describing orbit: transformations carry points *along* orbits. Accordingly, *any* differential in terms of coordinates describing orbit is an invariant differential.

The transformations affect only position on orbit. The transformation h applied to the position $[X]$ changes Euclidean volume (in R^L),

$$dh[X] = \left|\frac{\partial h[X]}{\partial [X]}\right| d[X],$$

by the positive-Jacobian factor

$$J_L(h:[X]) = \left|\frac{\partial h[X]}{\partial [X]}\right|.$$

The particular transformation $[X]$ applied to the reference value i changes volume by the factor

$$J_L([X]) = J_L([X]:i),$$

which can be used as a compensating factor to produce an *invariant differential*

$$d\mu([X]) = \frac{d[X]}{J_L([X])},$$

a measure of volume that is constant under any transformation in G:

$$d\mu(h[X]) = \frac{dh[X]}{J_L([hX])} = \frac{J_L(h:[X])\,d[X]}{J_L(h:[X])J_L([X])} = d\mu([X]).$$

The construction shows that it is the *unique* invariant differential that coincides with Euclidean volume at the identity i. (See Figure 6.) And the method of construction shows that any other invariant differential differs only by a constant of proportionality, the constant being the ratio of the differentials at the identity.

As an example consider further the positive affine group:

$$d[A, C][\bar{x}, s_x] = C^2\, d[\bar{x}, s_x],$$
$$J_2([A, C]:[\bar{x}, s_x]) = C^2,$$
$$J_2([\bar{x}, s_x]) = s_x^2,$$
$$d\mu([\bar{x}, s_x]) = \frac{d[\bar{x}, s_x]}{s_x^2} = \frac{d\bar{x}\, ds_x}{s_x^2}.$$

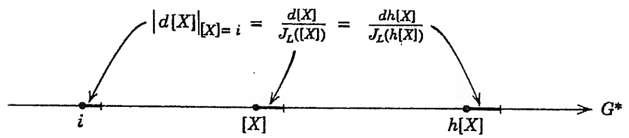


Figure 6 The invariant differential on G^* that coincides with $d[X]$ at the identity i.

Now consider the two invariant differentials as they apply to the element V and to images of V under transformations in G. At the reference point $D = D(X)$, let $\delta(D)$ be the ratio:

$$dm(X) = \frac{dX}{J_N(X)} = \delta(D) \frac{d[X]}{J_L([X])} = \delta(D)\, d\mu([X]).$$

The differentials, however, are invariant under transformations in G. The equality then holds throughout the orbit; the ratio $\delta(D)$ is a differential that measures V at right angles to the orbit.† This provides the change of variables from a volume element in the original coordinates to a volume element in terms of position coordinates conditional on a neighborhood of the orbit $D(X) = D$:

$$dm(X) = \delta(D)\, d\mu([X])$$

or

$$dX = \delta(D) \frac{J_N(X)}{J_L([X])} d[X].$$

As an example consider further the positive affine group:

$$dm(\mathbf{x}) = \delta(\mathbf{d})\, d\mu[\bar{x}, s_x],$$

$$d\mathbf{x} = \delta(\mathbf{d}) s_x^{n-2}\, d\bar{x}\, ds_x.$$

The factor $\delta(\mathbf{d})$ gives a measure of area on the sphere $s_x = 1$ in the subspace $\Sigma\, x_i = 0$; compare with Figure 15 in Chapter One. The variable $\sqrt{n}\,\bar{x}$ measures Euclidean distance in the direction of the $\mathbf{1}$-vector; the variable $\sqrt{n-1}\, s_x$ measures Euclidean distance orthogonal to the $\mathbf{1}$-vector (radially from the $\mathbf{1}$-vector): $\delta(\mathbf{d})/\sqrt{n}\sqrt{n-1}$ measures Euclidean area on the sphere $s_x = 1$ (with radius $\sqrt{n-1}$) in the subspace $\Sigma\, x_i = 0$.

4 THE ERROR PROBABILITY DISTRIBUTION

The conditional error density can now be derived. Assume that Assumption 3 in Section 3 holds and that the variable E has a density function $f(E)$ with respect to Euclidean volume on $\mathfrak{X}$.

The probability element for E is

$$\begin{aligned} f(E)\, dE &= f(E) J_N(E)\, dm(E) \\ &= \tilde{f}(E)\, dm(E); \end{aligned}$$

† The differential $\delta(D)$ can be written $\delta(D) = \delta_1(D)\,\delta_2$ where: (a) $\delta_2\, d\mu([X])$ at the identity gives L-dimensional Euclidean volume along orbit *as calculated using the coordinates of* R^N; and (b) $\delta_1(D)$ measures Euclidean volume at D in the $(N-L)$-dimensional space *orthogonal* to the orbit.

the modified $\tilde{f}(E)$ is a density with respect to the invariant differential $dm(E)$. The probability element can be expressed in terms of *orbit* $D = [E]^{-1}E$ and *position* $[E]$ by using results from the preceding section:

$$f([E]D)\, \delta(D) \frac{J_N(E)}{J_L([E])} d[E].$$

The *conditional probability element* is then obtained by normalization:

$$\begin{aligned} g([E]: D)\, d[E] &= k(D) f([E]D) \frac{J_N(E)}{J_L([E])} d[E] \\ &= \tilde{g}([E]: D)\, d\mu([E]) \\ &= k(D) \tilde{f}([E]D)\, d\mu([E]), \end{aligned}$$

where

$$k^{-1}(D) = \int_{G^*} f([E]D) \frac{J_N(E)}{J_L([E])} d[E].$$

The reduced structural model can now be given as

Reduced Structural Model

$$g([E]: D(X))\, d[E],$$

$$[X] = \theta[E].$$

The model has two parts: an *error probability distribution* $g([E]: D)\, d[E]$ ($[E]$ is a *variable* on G^*) which provides probability statements for the unknown $[E]$ in the structural equation; and a *structural equation* which gives the relation between the known $[X]$ in G^* and the unknowns θ in G and $[E]$ in G^* ($[E]$ is a *constant*).

Some distributions connected with the conditional error distribution need further results concerning invariant differentials. In Section 3 the transformation

$$\tilde{g} = hg$$

involving *left* multiplication by the group element h was examined. And the invariant differential, more correctly the *left invariant differential*, was derived:

$$d\mu(g) = \frac{dg}{J_L(g)}.$$

If the variable g is used as a position variable on an orbit, then the transformation $\tilde{g} = hg$ can be viewed as coming from a transformation $\tilde{X} = hX$ on the space $\mathfrak{X}$, and it gives position of a *new point* relative to a fixed reference point (Figure 7).

Now consider the transformation

$$\tilde{g} = gh$$

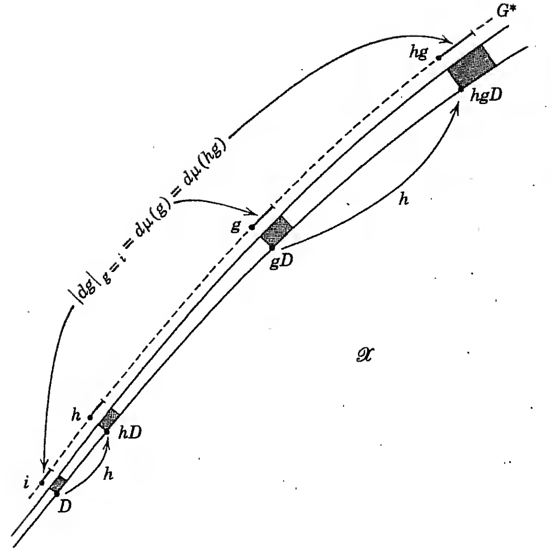


Figure 7 A left transformation: $\tilde{g} = hg$. The same volume measure at the new point as at the old point.

involving *right* multiplication by the group element h. If g is being used as a position variable on an orbit, then the "transformation" $\tilde{g} = gh$ can be viewed as a *change in reference point* from D to $h^{-1}D$ (see Figure 8). The right "transformation" h changes Euclidean volume,

$$dgh = \left|\frac{\partial gh}{\partial g}\right| dg,$$

by the positive-Jacobian factor

$$J_L^*(h:g) = \left|\frac{\partial gh}{\partial g}\right|.$$

A composite "transformation" h_1h_2 changes volume by the factor

$$J_L^*(h_1h_2:g) = J_L^*(h_2:gh_1)J_L^*(h_1:g).$$

The particular "transformation" g applied to the reference value i changes volume by the factor

$$J_L^*(g) = J_L^*(g:i),$$

which can be used as a compensating factor to produce the *right invariant differential*

$$d\nu(g) = \frac{dg}{J_L^*(g)},$$

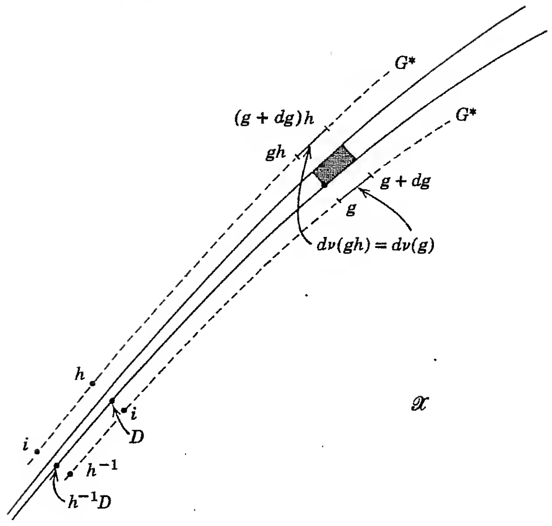


Figure 8 A right transformation: $g \to gh$. The same volume measure at a point using old or new coordinates.

a volume measure that is invariant under any right "transformation" $g \to gh$:

$$d\nu(gh) = \frac{dgh}{J_L^*(gh)} = \frac{J_L^*(h:g)\,dg}{J_L^*(h:g)J_L^*(g)} = d\nu(g).$$

The construction shows that it is the *unique* right invariant differential that coincides with Euclidean volume at the identity. *The right invariant differential measures volume invariantly under a change in reference point.*

As an example consider the positive affine group:

$$[\tilde{a}, \tilde{c}] = [a, c][A, C],$$
$$\tilde{a} = a + cA,$$
$$\tilde{c} = \quad cC,$$

$$J_2^*([A, C]:[a, c]) = \begin{vmatrix} 1 & A \\ 0 & C \end{vmatrix} = C,$$

$$J_2^*([a, c]) = c,$$

$$d\nu([a, c]) = \frac{da\,dc}{c}.$$

The left and right invariant differentials have some simple interconnecting properties. With g as variable the differential

$$d\mu_{g_0}(g) = d\mu(gg_0)$$

is left invariant:

$$d\mu_{g_0}(hg) = d\mu(hgg_0) = d\mu(gg_0) = d\mu_{g_0}(g).$$

Accordingly it must differ from the left invariant differential $d\mu(g)$ by a constant of proportionality, the ratio of volume measures at the identity:

$$d\mu(gg_0) = \Delta(g_0)\, d\mu(g).$$

The proportionality factor $\Delta(g)$ is called the *modular function* of the group. It has the following properties:

$$\begin{aligned}\Delta(i) &= 1,\\ \Delta(g_1g_2) &= \Delta(g_1)\,\Delta(g_2),\\ \Delta(g^{-1}) &= \Delta^{-1}(g).\end{aligned}$$

The factor $\Delta(g_0)$ measures a change under a right transformation; it can therefore be used as a compensating factor to construct a *right* invariant differential

$$d\nu_1(g) = \frac{d\mu(g)}{\Delta(g)} = \Delta^{-1}(g)\, d\mu(g).$$

The differential agrees with Euclidean volume at the identity; hence $d\nu_1(g) = d\nu(g)$.

A *right* invariant differential can also be constructed by assigning to a differential change at g the left invariant measure of the corresponding element at g^{-1}:

$$d\nu_2(g) = d\mu(g^{-1}).$$

This element is right invariant:

$$d\nu_2(gh) = d\mu(h^{-1}g^{-1}) = d\mu(g^{-1}) = d\nu_2(g).$$

At the identity i it agrees with Euclidean volume: Consider an element V at i; symmetrize the element,

$$V^s = \{g:\ \ g \in V \quad \text{or} \quad g^{-1} \in V\};$$

the μ measure and ν_2 measure of V^s are equal; hence $d\nu_2(g) = d\nu(g)$.

The invariant differentials can be interrelated:

$$\begin{aligned}d\mu(g) &= \Delta(g)\, d\nu(g) = d\nu(g^{-1}),\\ d\nu(g) &= \Delta^{-1}(g)\, d\mu(g) = d\mu(g^{-1}).\end{aligned}$$

The modular function has the form

$$\Delta(g) = \frac{J_L^*(g)}{J_L(g)}, \qquad \Delta(g^{-1}) = \frac{J_L(g)}{J_L^*(g)}.$$

For the positive affine group

$$\begin{aligned}\Delta([a, c]) &= \frac{J_2^*([a, c])}{J_2([a, c])} = \frac{c}{c^2} = \frac{1}{c},\\ d\mu([a, c]) &= \frac{da\,dc}{c^2} = \frac{1}{c}\,\frac{da\,dc}{c}\\ d\nu([a, c]) &= \frac{da\,dc}{c} = \left(\frac{1}{c}\right)^{-1}\frac{da\,dc}{c^2}.\end{aligned}$$

5 GENERAL INFERENCE

The structural model by its own information content produces the reduced model

$$\begin{aligned}[E]\!:\quad GE &= GX,\\ [X] &= \theta[E].\end{aligned}$$

Consider the information in this model concerning the unknown quantity θ.

The value of $[X]$ in the structural equation is known. Each possible value for the unknown $[E]$ corresponds to a possible value for θ:

$$\theta = [X][E]^{-1}, \qquad [E] = \theta^{-1}[X].$$

A probability statement concerning $[E]$ is *ipso facto* a probability statement concerning θ. The probability distribution describing the unknown $[E]$ thus gives a distribution, the *structural distribution*, that describes the unknown θ. The structural distribution is obtained from the error probability distribution of $[E]$ by the map

$$\theta = [X][E]^{-1}$$

from $[E]$ in G^* to θ in G.

Now suppose that Assumption 3 in Section 3 holds and that E has a density $f(E)$ with respect to Euclidean volume. The reduced model is

$$\begin{aligned}\bar{g}\big([E]\!:D(X)\big)\, d\mu([E]) &= k(D) f([E]D)\, J_N(E)\, d\mu([E])\\ [X] &= \theta[E].\end{aligned}$$

The structural distribution is obtained by substituting:

$$\begin{aligned}[E] &= \theta^{-1}[X],\\ d\mu([E]) &= \Delta([X])\, d\mu(\theta^{-1})\\ &= \Delta([X])\, d\nu(\theta)\\ &= \Delta(\theta^{-1}[X])\, d\mu(\theta).\end{aligned}$$

The structural probability element for θ on the space G is

$$\begin{aligned} g^*(\theta:X)\,d\theta &= \bar{g}(\theta^{-1}[X]:D)\,\Delta(\theta^{-1}[X])\,d\mu(\theta) \\ &= k(D)f(\theta^{-1}X)J_N(\theta^{-1}X)\,\Delta(\theta^{-1}[X])\,d\mu(\theta) \\ &= k([X]^{-1}X)f(\theta^{-1}X)J_N(\theta^{-1}X)\,\frac{J_L^*(\theta^{-1}[X])}{J_L(\theta^{-1}[X])}\frac{d\theta}{J_L(\theta)} \\ &= k([X]^{-1}X)f(\theta^{-1}X)J_N(\theta^{-1}X)\,\frac{J_L^*([X])}{J_L([X])}\frac{d\theta}{J_L(\theta)} \end{aligned}$$

As an example, the structural probability element for the measurement model can be obtained by substitution:

$$\begin{aligned} g^*([\mu,\sigma]:\mathbf{x})\,d\mu\,d\sigma &= k(\mathbf{d})\prod_1^n f([\mu,\sigma]^{-1}x_i)\left(\frac{s_x}{\sigma}\right)^n \frac{s_x/\sigma}{(s_x/\sigma)^2}\frac{d\mu\,d\sigma}{\sigma^2} \\ &= k(\mathbf{d})\prod_1^n f([\mu,\sigma]^{-1}x_i)\left(\frac{s_x}{\sigma}\right)^n \frac{1}{s_x}\frac{d\mu\,d\sigma}{\sigma}. \end{aligned}$$

6 TESTS OF SIGNIFICANCE

Consider the two tests of significance for the measurement model in Section 17 in Chapter One. The first test concerned the *hypothesis* $\mu = \mu_0$ and was based on the value of the error quantity $\bar{e}/s_e$. The possible hypotheses of the form $\mu = \mu_0$ produce a partition of G; and the possible values for the error quantity $\bar{e}/s_e$ produce a partition of G^* (see Figure 9).

The second test concerned the *hypothesis* $\sigma = \sigma_0$ and was based on the value of the error quantity s_e. The possible hypotheses of the form $\sigma = \sigma_0$ produce a partition of G; and the possible values of s_e produce a partition of G^* (see Figure 10).

Now consider the structural model and suppose that the space G is partitioned into disjoint sets; let $H(\theta)$ be the set containing θ (see Figure 11).

Consider the *hypothesis* $H(\theta) = H_0$. This hypothesis combined with the structural equation $[X] = \theta[E]$ or $[E] = \theta^{-1}[X]$ gives the information that the unknown $[E]$ is in the set

$$\begin{aligned} \{\theta^{-1}[X]:\ H(\theta) = H_0\} &= \{\theta^{-1}[X]:\ \theta \in H_0\} \\ &= H_0^{-1}[X] \end{aligned}$$

(note that $H^{-1} = \{g^{-1}:\ g \in H\}$ is the set of inverses of elements of H). The sets $H = H(\theta)$ form a partition of G; the corresponding sets $H^{-1}[X]$ form a partition P on G^*, a *reflection* about $[X]$ of the partition on G (see Figure 11).

The information concerning $[E]$ is that $[E]$ is in a set $H_0^{-1}[X]$, a set in the partition P of G^* into components $H^{-1}[X]$. Consider what the information

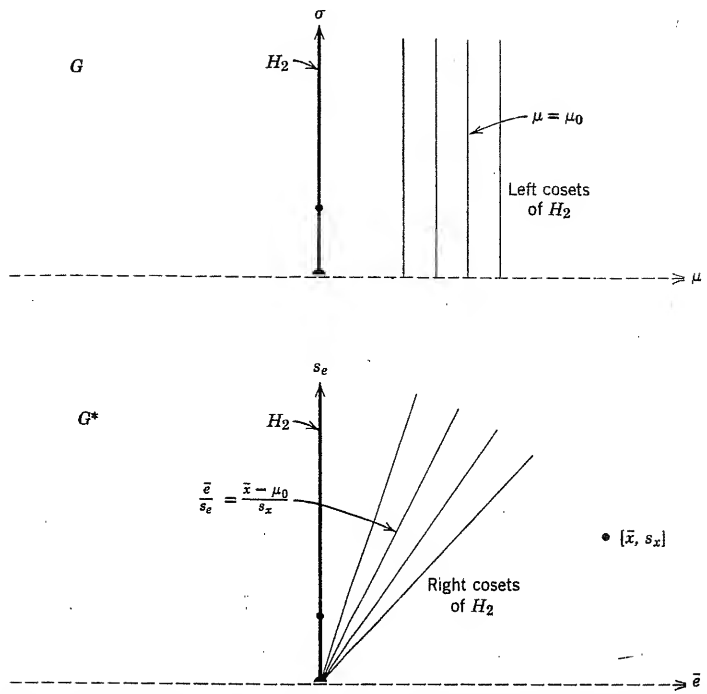


Figure 9 The *hypothesis* $\mu = \mu_0$ and a partition of G. The value of the error characteristic $\bar{e}/s_e$ and a partition of G^*.

concerning $[E]$ would be if $[X]$ were different. If $[X]$ were different, $g[X]$ for example, then the information would be that $[E]$ is in the set $H_0^{-1}g[X]$. If the presentation of information has the form of events, then the sets $H_0^{-1}g[X]$ (various g) must all be sets in the initial partition $P = \{H^{-1}[X]\}$ of G^*. This implies

$$\begin{aligned} H_0^{-1}g[X] &= H^{-1}[X], && \text{for some } H, \\ H_0^{-1}g &= H^{-1}, && \text{for some } H, \\ g^{-1}H_0 &= H, && \text{for some } H. \end{aligned}$$

Thus the left multiples gH_0 of H_0 must all be sets in the partition $\{H(\theta)\}$. By group theory it follows that *the partition* $\{H(\theta)\}$ *consists of left cosets* gH *of a subgroup* H *of* G. See Problems 12 and 13.

For the measurement model example, the *hypothesis* $\mu = \mu_0$ can be expressed as

$$[\mu, \sigma] \in [\mu_0, 1]H_2,$$

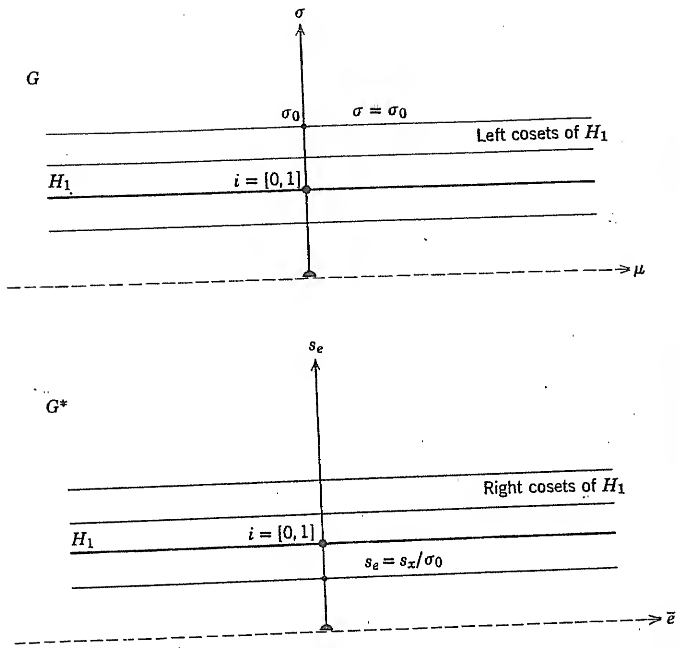


Figure 10 The *hypothesis* $\sigma = \sigma_0$ and a partition of G. The value of the error characteristic s_e and a partition of G^*.

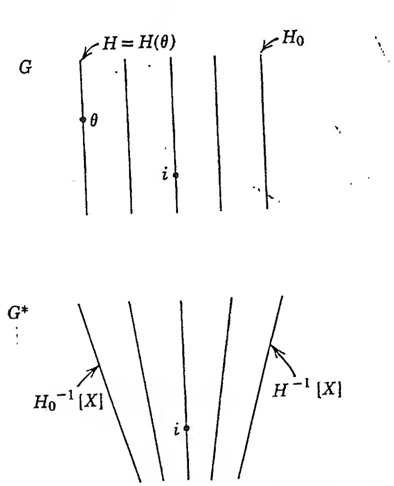


Figure 11 Sets H_0, $H = H(\theta)$ in the partition $\{H(\theta)\colon\ \theta \in G\}$ of G. The corresponding sets $H_0^{-1}[X]$, $H^{-1}[X]$ in the partition $P = \{H^{-1}(\theta)[X]\colon\ \theta \in G\}$ of G^*.

a left coset of the scale group H_2. Show that the *hypothesis* $\sigma = \sigma_0$ can also be put in left coset form.

Suppose now that some outside source has indicated that θ is in the *left coset* g_0H of a subgroup H of G. The *hypothesis* $\theta \in g_0H$ gives information concerning the unknown error $[E]$ in the structural equation: The error $[E]$ is in the set

$$\begin{aligned}\{\theta^{-1}[X]\colon\ \theta \in g_0H\} &= (g_0H)^{-1}[X]\\ &= H^{-1}g_0^{-1}[X]\\ &= Hg_0^{-1}[X],\end{aligned}$$

which is a *right coset* or orbit on G^*. (Note that $H^{-1} = H$ since H is a group.) Thus the *hypothesis* $\theta \in g_0H$ (a left coset on G) gives the information that $[E]$ is on the orbit $Hg_0^{-1}[X]$ (a right coset on G^*). This information has the form

$$H[E] = Hg_0^{-1}[X];$$

it is an event for the error variable and it uses the orbital variable $H[E]$ on G^* or HE on $\mathfrak{X}$.

Let $t([E])$ be a variable that indexes the H orbits on G^*. The *hypothesis* $\theta \in g_0H$ together with the structural equation leads to the value

$$t([E]) = t(g_0^{-1}[X]).$$

This value can be compared with the distribution of the variable $t([E])$ derived from the error probability distribution; and the hypothesis can be assessed accordingly.

For the measurement model example the information that $\bar{e}/s_e$ is equal to

$$\frac{\bar{x} - \mu_0}{s_x}$$

is equivalent to the information that $[\bar{e}, s_e]$ is on the orbit

$$H_2[\mu_0, 1]^{-1}[\bar{x}, s_x] = H_2[\bar{x} - \mu_0, s_x],$$

a right coset of the scale group H_2 (see Figure 9).

*7 CONDITIONING BY OUTSIDE INFORMATION

The measurement model with normal error is a simple example to illustrate conditioning. In reduced form the model is

$$\bar{e} = \frac{z}{\sqrt{n}}, \qquad s_e = \frac{\chi_{n-1}}{\sqrt{n-1}},$$

$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e]$$

(Section 16 in Chapter One). The corresponding structural distribution is

$$\mu = \bar{x} - s_x \frac{\bar{e}}{s_e},$$

$$\sigma = s_x \frac{1}{s_e},$$

where $[\bar{e}, s_e]$ has the error probability distribution (Section 18 in Chapter One).

Suppose the information $\sigma = \sigma_0$ becomes available. This information has the form of an event for the error variable (Section 6); it leads to the value

$$s_e = \frac{s_x}{\sigma_0}$$

for the error characteristic

$$s_e = \frac{\chi_{n-1}}{\sqrt{n-1}}.$$

As an event it can be used to condition the error distribution; it gives

$$\bar{e} = \frac{z}{\sqrt{n}}, \qquad s_e = \frac{s_x}{\sigma_0},$$

which then gives

$$\mu = \bar{x} - s_x \frac{z/\sqrt{n}}{s_x/\sigma_0} = \bar{x} - z\frac{\sigma_0}{\sqrt{n}},$$

$$\sigma = s_x \frac{1}{s_x/\sigma_0} = \sigma_0$$

for the structural distribution. This structural distribution for μ, σ *given the information* $\sigma = \sigma_0$ is the same as the structural distribution from the simple measurement model

$$\bar{e} = \frac{\sigma_0}{\sqrt{n}} z,$$

$$\bar{x} = \mu + \bar{e}$$

in Sections 8, 10 of Chapter One.

In this example, outside information concerning $[\mu, \sigma]$ was introduced. The information had the form of an event for the error variable, and it produced a conditioned error distribution and a conditioned structural distribution. These conditioned distributions are the *same* as those that would have been obtained had the information been available when the model was constructed. *This agreement holds for the general structural model.*

Consider the structural model in reduced form:

$$[E]{:}\,GE = GX,$$

$$[X] = \theta[E],$$

and suppose there is information concerning θ in the form of an event. By Section 6 the information has the form $\theta \in g_0 H_1$, where H_1 is a subgroup of G.

In the typical case involving a subgroup H_1 in a group G there is a complementary subgroup H_2 such that each element g in G can be written *uniquely* as a product:

$$g = kh,$$

where k is in H_2 and h is in H_1. This kind of decomposition† for G leads to convenient notation; it is examined here in lieu of the general case. Suppose G can be expressed in this manner, $g = kh$, and let

$$\underset{2}{[g]} = k, \qquad \underset{1}{[g]} = h,$$

$$g = \underset{2}{[g]}\underset{1}{[g]}.$$

The inverse of an element g is

$$g^{-1} = \underset{1}{[g]}^{-1}\underset{2}{[g]}^{-1}.$$

An inverse can be any element in G; accordingly the decomposition of G can be made in the reverse order:

$$g = \underset{1}{[g]}\underset{2}{[g]}$$

and

$$\underset{1}{[g^{-1}]} = \underset{1}{[g]}^{-1}, \qquad \underset{2}{[g^{-1}]} = \underset{2}{[g]}^{-1}.$$

(See Figure 12.)

The quantity θ in G can be represented in terms of elements of H_2 and H_1:

$$\theta = \tau\varphi, \qquad \tau = \underset{2}{[\theta]}, \qquad \varphi = \underset{1}{[\theta]}.$$

The information $\theta \in g_0 H_1$ can then be given as $\tau = \tau_0$, where $\tau_0 = \underset{2}{[g_0]}$ and

$$g_0 H_1 = \underset{2}{[g_0]}\underset{1}{[g_0]} H_1 = \tau_0 H_1.$$

(See Figure 13.)

† G is called the *semidirect product* of the subgroups H_2 and H_1.

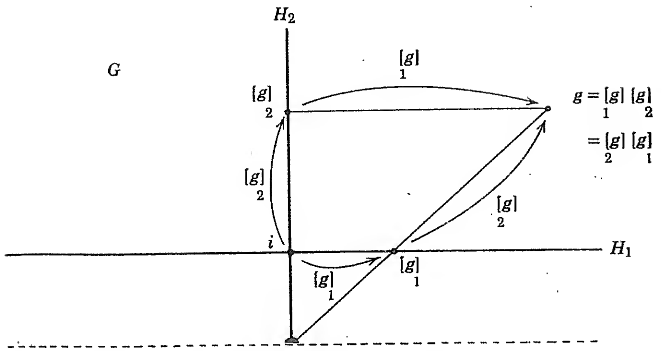


Figure 12 G expressed as a semidirect product of H_2 and H_1 and as a semidirect product of H_1 and H_2.

The information $\tau = \tau_0$ gives information concerning the unknown error position $[E]$. The structural equation in the form $[E] = \theta^{-1}[X]$ can be written

$$\underset{1}{[E]}\underset{2}{[E]} = \varphi^{-1}\tau_0^{-1}[X] = \varphi^{-1}[\tau_0^{-1}X]$$

and then separated:

$$\underset{1}{[E]} = \varphi^{-1}\underset{1}{[\tau_0^{-1}X]},$$

$$\underset{2}{[E]} = \underset{2}{[\tau_0^{-1}X]}.$$

The information $\tau = \tau_0$ determines the H_1 orbit of $[E]$ on G^* (also the H_1

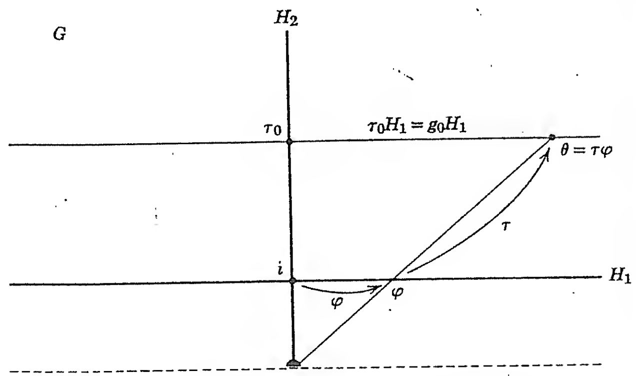


Figure 13 The information $\theta \in \tau_0 H_1$ or, equivalently, $\tau = \tau_0$.

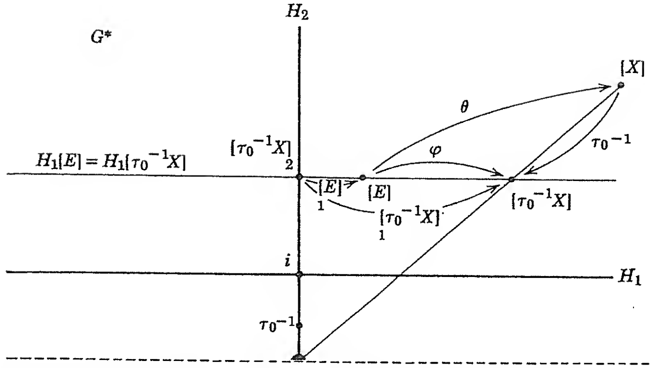


Figure 14 The orbit of $[E]$ under H_1 as determined by the information $\tau = \tau_0$.

orbit of E on $\mathfrak{X}$); the orbit can be designated alternatively by the reference point

$$\underset{1}{[E]}^{-1}[E] = \underset{2}{[E]} = \underset{2}{[\tau_0^{-1}X]}$$

on G^*. The information also produces a restricted structural equation describing position on that orbit:

$$\underset{1}{[E]} = \varphi^{-1}\underset{1}{[\tau_0^{-1}X]}, \qquad \underset{1}{[\tau_0^{-1}X]} = \varphi\underset{1}{[E]}.$$

(See Figure 14.)

The error distribution conditional on its H_1 orbit can be expressed as

$$\underset{1}{[E]}:\quad H_1[E] = H_1[\tau_0^{-1}X], \qquad GE = GX,$$

or equivalently as

$$\underset{1}{[E]}:\quad H_1E = H_1\tau_0^{-1}X.$$

The conditioned model then has the form

Conditioned Model

$$\underset{1}{[E]}:\quad H_1E = H_1\tau_0^{-1}X,$$

$$\underset{1}{[\tau_0^{-1}X]} = \varphi\underset{1}{[E]}.$$

Alternatively, suppose the information $\tau = \tau_0$ had been available when the model was being constructed. The error distribution would be

$$E.$$

And the structural equation would be

$$X = \tau_0 \varphi E$$

or

$$Y = \tau_0^{-1} X = \varphi E.$$

This would give the structural model

$$E,$$
$$Y = \varphi E$$

based on a quantity φ in the group H_1. The reduced model would be

$$\underset{1}{[E]}: \quad H_1 E = H_1 Y,$$

$$\underset{1}{[Y]} = \varphi \underset{1}{[E]}.$$

This is the same as the conditioned model in the preceding paragraph.

*8 MARGINAL AND CONDITIONAL DISTRIBUTIONS

The test of significance in Section 6 required a *marginal* distribution of an error variable. The use of outside information in Section 7 produced a *conditioned* distribution of the error variable. Marginal and conditional distributions can be derived quite generally. They provide a *decomposition* of the error distribution and a corresponding decomposition of the structural distribution.

Consider the general structural model with error density $f(E)$ on $\mathfrak{X}$; suppose that Assumption 3 (Section 3) holds and that the quantity θ in G can be factored uniquely:

$$\theta = \tau \varphi$$

where τ and φ are in subgroups H_2 and H_1, respectively.

Let $[X]$ be a transformation variable with invariant differential

$$dm(X) = \frac{dX}{J_N(X)}$$

on $\mathfrak{X}$. Let $d\mu(g)$, $d\nu(g)$ with $\Delta(g)$ be invariant differentials on G, $d\mu_1(h)$, $d\nu_1(h)$ with $\Delta_1(h)$ be invariant differentials on H_1 (an open set in L_1 dimensions), and $d\mu_2(k)$, $d\nu_2(k)$ with $\Delta_2(k)$ be invariant differentials on H_2 (an open set in L_2 dimensions).

The adjusted differential

$$\frac{d\mu(hk)}{\Delta(k)}$$

is invariant under multiplication on the left by an element of H_1 and on the right by an element of H_2. The composite differential

$$d\mu_1(h)\,\frac{d\mu_2(k)}{\Delta_2(k)}$$

has the same invariance property. Let δ be their ratio at the identity i then

$$d\mu(hk) = \delta\, d\mu_1(h)\,\frac{\Delta(k)}{\Delta_2(k)}\,d\mu_2(k) = \delta\, d\mu_1(h)\,\Delta(k)\,d\nu_2(k).$$

The substitution $h \to h^{-1}$, $k \to k^{-1}$ leads to a parallel relation:

$$d\nu(kh) = \delta\, d\nu_1(h)\,\frac{\Delta^{-1}(k)}{\Delta_2^{-1}(k)}\,d\nu_2(k) = \delta\, d\nu_1(h)\,\Delta^{-1}(k)\,d\mu_2(k).$$

The probability element for the error variable $[E]$ can be expressed in terms of the components $\underset{1}{[E]}$ and $\underset{2}{[E]}$:

$$\begin{aligned}
\bar{g}([E]:D)\,d\mu([E]) &= k(D)\bar{f}([E]D)\,d\mu([E]) \\
&= k(D)\bar{f}(\underset{1}{[E]}\underset{2}{[E]}D)\,\delta\,d\mu_1(\underset{1}{[E]})\,\frac{\Delta(\underset{2}{[E]})}{\Delta_2(\underset{2}{[E]})}\,d\mu_2(\underset{2}{[E]}) \\
&= k_1(\underset{2}{[E]}D)\bar{f}(\underset{1}{[E]}\underset{2}{[E]}D)\,d\mu_1(\underset{1}{[E]}) \cdot \frac{k(D)\,\delta}{k_1(\underset{2}{[E]}D)}\,\frac{\Delta(\underset{2}{[E]})}{\Delta_2(\underset{2}{[E]})}\,d\mu_2(\underset{2}{[E]}).
\end{aligned}$$

The last expression is written as a product of the conditional distribution of $\underset{1}{[E]}$ given $\underset{2}{[E]}$ and the marginal distribution of $\underset{2}{[E]}$; the constant $k_1(\underset{2}{[E]}D)$ normalizes the conditional distribution.

The structural probability element for $\theta = \tau\varphi$ can similarly be expressed in terms of the components φ in H_1 and τ in H_2:

$$\begin{aligned}
g^*(\theta : X)\,d\theta &= k(D)\bar{f}(\theta^{-1}X)\,\Delta([X])\,d\nu(\theta) \\
&= k(D)\bar{f}(\varphi^{-1}\tau^{-1}X)\,\Delta([X])\,\delta\,d\nu_1(\varphi)\,\Delta^{-1}(\tau)\,d\mu_2(\tau).
\end{aligned}$$

The restricted structural equation for φ and $\underset{1}{[E]}$ is

$$\underset{1}{[E]} = \varphi^{-1}\underset{1}{[\tau^{-1}X]},$$

and the corresponding differential is

$$d\mu_1([E]_1) = \Delta_1([\tau^{-1}X]_1)\, d\nu_1(\varphi);$$

the remainder of the structural equation is

$$[E]_2 = [\tau^{-1}X]_2.$$

The normalizing constant of the conditional error distribution can then be used:

$$g^*(\theta : X)\, d\theta = k_1([\tau^{-1}X]_2 D) f(\varphi^{-1}[\tau^{-1}X]D)\, \Delta_1([\tau^{-1}X]_1)\, d\nu_1(\varphi)$$
$$\cdot \frac{k(D)\,\delta}{k_1([\tau^{-1}X]_2 D)} \frac{\Delta(\tau^{-1}[X])}{\Delta_1([\tau^{-1}X]_1)}\, d\mu_2(\tau).$$

This expresses the structural distribution as a product of the *conditional distribution of φ given τ*, and *the marginal distribution of τ*.

NOTES AND REFERENCES

In this chapter the essential elements of the measurement model have been placed in a general framework. The resulting model, the structural model, covers a broad range of problems, many of which are examined in succeeding chapters.

Some aspects of the structural model can be found in Fraser (1961). The main pattern of development here appears in Fraser (1966, 1967).

The use of transformation groups (for example, James, 1954) may be found recurrently through the statistical literature, always with respect to the classical model and usually as a device to gain simplicity. In contrast, their use here is primary and essential: the *second* ingredient of the model.

The analysis of conditional and marginal error distributions and of factorizations of the invariant measures has developed in interchange with A. Kalotay, H. Levenbach, and J. Whitney.

Fraser D. A. S. (1961), The fiducial method and invariance, *Biometrika*, **48**, 261–280.
Fraser D. A. S. (1966), Structural probability and a generalization, *Biometrika*, **53**, 1–9.
Fraser D. A. S. (1967), Data transformations and the linear model, *Ann. Math. Statistics*, **38**, 1456–1465.
James A. T. (1954), Normal multivariate analysis and the orthogonal group, *Ann. Math. Statistics*, **25**, 40–75.
Lehmann E. L. (1959), *Testing Statistical Hypotheses*, Wiley, New York.
Peisakoff M. (1951), Transformation parameters, Ph.D. thesis, Princeton University.

PROBLEMS

1. A group G is called *commutative* or *Abelian* if $gh = hg$ for all g, h in G.

(i) Show that the location group and the scale group are Abelian.
(ii) Show that the positive affine group is non-Abelian.
(iii) For the location group

$$H_1 = \{[a, 1]: \quad -\infty < a < \infty\}$$

show that

$$d\mu([a, 1]) = da, \qquad d\nu([a, 1]) = da,$$
$$\Delta([a, 1]) \equiv 1.$$

(iv) For the scale group

$$H_2 = \{[0, c]: \quad 0 < c < \infty\}$$

show that

$$d\mu([0, c]) = \frac{dc}{c} = d\ln c, \qquad d\nu([0, c]) = \frac{dc}{c} = d\ln c,$$
$$\Delta([0, c]) \equiv 1.$$

(v) For an Abelian group G show that $\Delta(g) = 1$ for all g in G.

2. Consider the *location group*

$$G_1 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ a_1 & 1 & 0 \\ a_2 & 0 & 1 \end{pmatrix} : \begin{array}{c} -\infty < a_1 < \infty \\ -\infty < a_2 < \infty \end{array} \right\}$$

acting on R^2:

$$\tilde{\mathbf{x}} = \begin{pmatrix} 1 \\ \tilde{x} \\ \tilde{x} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ a_1 & 1 & 0 \\ a_2 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = g\mathbf{x}.$$

The extra first element in the vector permits the use of matrix multiplication.

(i) Show that G_1 is an Abelian group.
(ii) Show that G_1 is unitary on R^2.
(iii) Show that $G\mathbf{x} = R^2$.
(iv) Show that

$$dm(\mathbf{x}) = dx_1\, dx_2,$$
$$d\mu(g) = da_1\, da_2, \qquad d\nu(g) = da_1\, da_2,$$
$$\Delta(g) \equiv 1.$$

(v) Show that g can be expressed alternatively as

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]$$

using the notation of Problem 27, Chapter One. Give the formula for

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

3. (*Continuation*). Consider the preceding group acting on R^{2n}:

$$\tilde{X} = \begin{pmatrix} 1 & \cdots & 1 \\ \tilde{x}_{11} & \cdots & \tilde{x}_{1n} \\ \tilde{x}_{21} & \cdots & \tilde{x}_{2n} \end{pmatrix} = g\begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \end{pmatrix} = gX.$$

(i) Show that the group is unitary on R^{2n}.
(ii) Determine the form of the orbits GX. *Note: X* can be viewed as a point $(x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n})$ in R^{2n} or as an ordered pair of points, $(x_{11}, \ldots, x_{1n})$, $(x_{21}, \ldots, x_{2n})$, in R^n.
(iii) Show that

$$[X] = \begin{pmatrix} 1 & 0 & 0 \\ \bar{x}_1 & 1 & 0 \\ \bar{x}_2 & 0 & 1 \end{pmatrix}$$

is a transformation variable. Give three other examples of transformation variables.
(iv) Show that

$$dm(X) = \prod_{i=1}^{n} (dx_{1i}\, dx_{2i}).$$

(v) Give the formula for

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \end{pmatrix}.$$

4. Consider the *scale group*

$$G_2 = \left\{\begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & 0 & c_2 \end{pmatrix} : \begin{matrix} 0 < c_1 < \infty \\ 0 < c_2 < \infty \end{matrix}\right\}$$

acting on R^{2n}:

$$\tilde{X} = \begin{pmatrix} 1 & \cdots & 1 \\ \tilde{x}_{11} & \cdots & \tilde{x}_{1n} \\ \tilde{x}_{21} & \cdots & \tilde{x}_{2n} \end{pmatrix} = g\begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \end{pmatrix} = gX.$$

(i) Show that G_2 is an Abelian group.
(ii) Show that the group is unitary on R^{2n}; omit points X having $(x_{11}, \ldots, x_{1n}) = 0'$ or $(x_{21}, \ldots, x_{2n}) = 0'$.
(iii) Determine the form of the orbits G_2X; see the *Note* in Problem 3.
(iv) Show that

$$[X] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\sum x_{1i}^2)^{1/2} & 0 \\ 0 & 0 & (\sum x_{2i}^2)^{1/2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & s_1(X) & 0 \\ 0 & 0 & s_2(X) \end{pmatrix}$$

is a transformation variable.

(v) Show that

$$dm(X) = \frac{\prod (dx_{1i}\, dx_{2i})}{s_1^n(X)s_2^n(X)}$$

$$d\mu(g) = \frac{dc_1\, dc_2}{c_1c_2}, \qquad d\nu(g) = \frac{dc_1\, dc_2}{c_1c_2},$$

$$\Delta(g) \equiv 1.$$

(vi) Show that g can be expressed alternatively as

$$\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}\right].$$

Give the formula for

$$\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix}\right]\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \end{pmatrix}.$$

5. Let $d\mu(g)$ be the left invariant differential on G that agrees with Euclidean volume at the identity: Assumption 3 in Section 3. Show that the *compensated* differential

$$d\nu_1(g) = \Delta^{-1}(g)\, d\mu(g)$$

is right invariant and agrees with Euclidean volume at the identity (see Section 4).

6. Consider the *shear group*

$$G_3 = \left\{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 0 \end{pmatrix}, \quad -\infty < k < \infty\right\}$$

acting on R^2:

$$\tilde{\mathbf{x}} = \begin{pmatrix} 1 \\ \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{pmatrix}\begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} = g\mathbf{x}.$$

(i) Show that G_3 is an Abelian group.
(ii) Show that G_3 is unitary on R^2; omit the points having $x_1 = 0$.
(iii) Find the form of the orbits $G_3\mathbf{x}$.
(iv) Show that

$$[\mathbf{x}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & x_2x_1^{-1} & 1 \end{pmatrix}$$

is a transformation variable.
(v) Show that

$$dm(\mathbf{x}) = dx_1\, dx_2,$$

$$d\mu(g) = dk, \qquad d\nu(g) = dk,$$

$$\Delta(g) \equiv 1.$$

(vi) Show that g can be expressed alternatively as

$$\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}\right].$$

Give the formula for

$$\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & 0\\ k & 1\end{pmatrix}\right]\begin{pmatrix}x_1\\ x_2\end{pmatrix}.$$

7 (*Continuation*). Consider the group in Problem 6 acting on R^{2n}:

$$\tilde{X} = \begin{pmatrix}1 & \cdots & 1\\ \tilde{x}_{11} & \cdots & \tilde{x}_{1n}\\ \tilde{x}_{21} & \cdots & \tilde{x}_{2n}\end{pmatrix} = g\begin{pmatrix}1 & \cdots & 1\\ x_{11} & \cdots & x_{1n}\\ x_{21} & \cdots & x_{2n}\end{pmatrix} = gX.$$

(i) Show that the group is unitary on R^{2n}; omit points with $(x_{11}, \ldots, x_{1n}) = 0'$.
(ii) Show that

$$[X] = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & t(X) & 1\end{pmatrix}$$

where $t(X) = \sum x_{1i}x_{2i}/\sum x_{1i}^2$ is a transformation variable.
(iii) Show that

$$dm(X) = \prod (dx_{1i}\, dx_{2i}).$$

(iv) Give the formula for

$$\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & 0\\ k & 1\end{pmatrix}\right]\begin{pmatrix}x_{11} & \cdots & x_{1n}\\ x_{21} & \cdots & x_{2n}\end{pmatrix}.$$

8. Consider the *progression group* (scale and shear group)

$$G_4 = \left\{\begin{pmatrix}1 & 0 & 0\\ 0 & c_1 & 0\\ 0 & k & c_2\end{pmatrix} : \begin{matrix}0 < c_j < \infty\\ -\infty < k < \infty\end{matrix}\right\}$$

acting on R^{2n} (X as in problem 7):

$$\tilde{X} = gX.$$

(i) Show that G_4 is a group.
(ii) For $n \geq 2$ show that G_4 is unitary on R^{2n}; omit points having $(x_{11}, \ldots, x_{1n})$ and $(x_{21}, \ldots, x_{2n})$ linearly dependent.
(iii) Describe an orbit GX; for convenience represent X as an ordered pair of points in R^n.
(iv) Show that

$$[X] = \begin{pmatrix}1 & 0 & 0\\ 0 & s_1(X) & 0\\ 0 & t(X) & s_{(2)}(X)\end{pmatrix}$$

is a transformation variable, where $s_1(X) = (\sum x_{1i}^2)^{1/2}$,

$$t(X) = \frac{\sum x_{1i}x_{2i}}{s_1(X)}, \qquad s_{(2)}(X) = [\sum (x_{2i} - t(X)x_{1i}/s(X))^2]^{1/2}.$$

(v) Show that

$$dm(X) = \frac{\prod (dx_{1i}\, dx_{2i})}{s_1^n(X)s_{(2)}^n(X)},$$

$$d\mu(g) = \frac{dc_1\, dk\, dc_2}{c_1c_2^2}, \qquad d\nu(g) = \frac{dc_1\, dk\, dc_2}{c_1^2c_2},$$

$$\Delta(g) = \frac{c_1}{c_2}.$$

(vi) Show that g can be expressed alternatively as

$$\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}c_1 & 0\\ k & c_2\end{pmatrix}\right];$$

check the multiplication. Give the formula for

$$\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}c_1 & 0\\ k & c_2\end{pmatrix}\right]\begin{pmatrix}x_{11} & \cdots & x_{1n}\\ x_{21} & \cdots & x_{2n}\end{pmatrix}.$$

9. Consider the *location-progression group*

$$G_5 = \left\{\begin{pmatrix}1 & 0 & 0\\ a_1 & c_1 & 0\\ a_2 & k & c_2\end{pmatrix} : \begin{matrix}0 < c_j < \infty\\ -\infty < k < \infty\\ -\infty < a_j < \infty\end{matrix}\right\}$$

acting on R^{2n} (X as in problem 7):

$$\tilde{X} = gX.$$

(i) Show that G_5 is a group.
(ii) For $n \geq 3$ show that G_5 is unitary on R^{2n}; omit points with $(1, \ldots, 1)$, $(x_{11}, \ldots, x_{1n})$, $(x_{21}, \ldots, x_{2n})$ linearly dependent.
(iii) Show that

$$[X] = \begin{pmatrix}1 & 0 & 0\\ \bar{x}_1 & s_1(X) & 0\\ \bar{x}_2 & t(X) & s_{(2)}(X)\end{pmatrix}$$

is a transformation variable with

$$s_1(X) = [\sum (x_{1i} - \bar{x}_1)^2]^{1/2},$$

$$t(X) = \frac{\sum (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{s_1(X)},$$

$$s_{(2)}(X) = [\sum (x_{2i} - \bar{x}_2 - t(X)(x_{1i} - \bar{x}_1)/s_1(X))^2]^{1/2}.$$

(iv) Show that

$$dm(X) = \frac{\prod (dx_{1i}\, dx_{2i})}{s_1^n(X)s_{(2)}^n(X)},$$

$$d\mu(g) = \frac{da_1\, da_2\, dc_1\, dk\, dc_2}{c_1^2c_2^3},$$

$$d\nu(g) = \frac{da_1\, da_2\, dc_1\, dk\, dc_2}{c_1^2c_2^1},$$

$$\Delta(g) = \frac{1}{c_2^2}.$$

(v) Show that g can be expressed alternatively as

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right].$$

Give the formula for

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right]\begin{pmatrix} x_{11} & \cdots & x_{2n} \\ x_{21} & \cdots & x_{2n} \end{pmatrix}.$$

10. Consider the *positive linear group*

$$G_6 = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{11} & c_{12} \\ 0 & c_{21} & c_{22} \end{bmatrix} : \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} > 0 \right\}$$

acting on R^{2n} (X as in Problem 7):

$$\tilde{X} = gX.$$

(i) Show that G_6 is a group, a non-Abelian group.

(ii) For $n \geq 2$ show that G_6 is unitary on R^{2n}; omit the points X having $(x_{11}, \ldots, x_{1n})$ and $(x_{21}, \ldots, x_{2n})$ linearly dependent.

(iii) Show that

$$d\mu(g) = \frac{dc_{11}\, dc_{21}\, dc_{12}\, dc_{22}}{|g|^2},$$

$$d\nu(g) = \frac{dc_{11}\, dc_{12}\, dc_{21}\, dc_{22}}{|g|^2},$$

$$\Delta(g) \equiv 1.$$

(iv) Show that *an* invariant differential is

$$dm(X) = \frac{\prod (dx_{1i}\, dx_{2i})}{|XX'|^{n/2}}.$$

*(v) Develop a transformation variable.

11. Consider the *positive affine group* (location positive-linear group)

$$G_7 = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ a_1 & c_{11} & c_{12} \\ a_2 & c_{21} & c_{22} \end{bmatrix} : \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} > 0 \right\}$$

acting on R^{2n} (X as in Problem 7):

$$\tilde{X} = gX.$$

(i) Show that G_7 is a group.

(ii) For $n \geq 3$ show that G_7 is unitary on R^{2n}; omit points X having $(1, \ldots, 1)$, $(x_{11}, \ldots, x_{1n})$, $(x_{21}, \ldots, x_{2n})$ linearly dependent.

(iii) Show that

$$d\mu(g) = \frac{da_1\, da_2\, dc_{11}\, dc_{21}\, dc_{12}\, dc_{22}}{|g|^3},$$

$$d\nu(g) = \frac{da_1\, dc_{11}\, dc_{12}\, da_2\, dc_{21}\, dc_{22}}{|g|^2},$$

$$\Delta(g) = |g|^{-1}.$$

(iv) Show that *an* invariant differential is

$$dm(X) = \frac{\prod (dx_{1i}\, dx_{2i})}{|XX'|^{n/2}}.$$

*(v) Develop a transformation variable.

12. Let H be a subgroup of a group G.

(i) Show that the sets $\{gH:\ g \in G\}$ form a partition of G, *the left cosets of the subgroup* H (*Example:* The first part of Figure 9; G is the positive affine group, and H is the scale group). Similarly, show that the sets $\{Hg:\ g \in G\}$ form a partition of G, *the right cosets of* H (*Example:* The second part of Figure 9). The set-forming braces are used in the free sense: the set of *distinct* entities gH formed as g takes values in G (the set $g_1H = g_2H$ with $g_1 \neq g_2$ occurs *once* in $\{gH:\ g \in H\}$).

(ii) H is a *normal subgroup in* G if $gH = Hg$ for all g in G. Show that the *partition* into left cosets is the same as the *partition* into right cosets if and only if H is normal in G. (*Example:* Figure 10; G is the positive affine group and H is the location group.)

*13. Consider a group G and a partition $\{H_\alpha\}$ of G; suppose that the partition $\{H_\alpha\}$ is closed under left multiplication by any element in G (i.e., for any g and H_α there is a set H_β in the partition such that $gH_\alpha = H_\beta$). Show that one of the sets H_α is a subgroup H of G and that the partition is by left cosets of H.

14. Consider a partition $\{gH\}$ of a group G into cosets with respect to a normal subgroup H.

(i) Show that a natural multiplication of cosets is *defined* by

$$(g_1H)(g_2H) = g_1g_2H.$$

(ii) Show that the multiplication rule for cosets satisfies the axioms of a group. This group defined on the cosets is the *factor group* G/H of G by the normal subgroup H.

15. Consider the notation for semidirect products on p. 69 in Section 7. Show that

$$\underset{1}{[g]}^{-1} = \underset{1}{[g^{-1}]}, \qquad \underset{2}{[g]}^{-1} = \underset{2}{[g^{-1}]}.$$

16. (i) Show that the location group H_1 is a normal subgroup of the positive affine group G (Section 11 in Chapter One and Figure 8 in Chapter One).

(ii) Show that the factor group G/H_1 can be represented by the scale group H_2 (Figure 8 in Chapter One).

(iii) Show that

$$[a, c] = \underset{1}{[a, c]}\underset{2}{[a, c]} = [a, 1][0, c]$$

$$= \underset{2}{[a, c]}\underset{1}{[a, c]} = [0, c][c^{-1}a, 1].$$

Note that

$$\underset{2}{[a, c]} = [0, c],$$

$$\underset{2}{[a, c]} = [0, c],$$

a consequence of normality of H_1 (see Figure 12 with G, H_1, H_2 as defined here).

*17. Consider the example at the beginning of Section 6. Use the notation of Section 7 and the results in Problem 16.

(i) Examine the hypothesis $\mu = \mu_0$ or $[\mu, \sigma] \in [\mu_0, 1]H_2$ (Figure 9). Show that the orbits of H_2 on G^* can be indexed by $t([\tilde{e}, s_e]) = \underset{1}{[\tilde{e}, s_e]}$ (i.e., show that $\underset{1}{[\tilde{e}, s_e]} = [\tilde{e}/s_e, 1]$ is in one-to-one correspondence with $H_2[\tilde{e}, s_e]$). Determine the value of

$$t([\mu_0, 1]^{-1}[\tilde{x}, s_x]).$$

Compare with Section 17 in Chapter One.

(ii) Examine hypothesis $\sigma = \sigma_0$ or $[\mu, \sigma] \in [0, \sigma_0]H_1$ (Figure 10). Show that the orbits of H_1 on G^* can be indexed by $t([\tilde{e}, s_e]) = \underset{2}{[\tilde{e}, s_e]}$ (i.e., show that $\underset{2}{[\tilde{e}, s_e]}$ is in one-to-one correspondence with $H_1[\tilde{e}, s_e]$). Determine the value of

$$t([0, \sigma_0]^{-1}[\tilde{x}, s_x]).$$

Compare with Section 17 in Chapter One.

18. (i) Show that the location group G_1 (Problem 2) is a normal subgroup of the location-progression group G_5 (Problem 9).

(ii) Show that the factor group G_5/G_1 can be represented by the progression group G_4 (Problem 8).

(iii) Show that

$$\underset{4}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & k & c_2 \end{pmatrix};$$

$$\underset{4}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & k & c_2 \end{pmatrix},$$

a consequence of normality of H_1.

(iv) Show that

$$\underset{1}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ a_1 & 1 & 0 \\ a_2 & 0 & 1 \end{pmatrix},$$

$$\underset{1}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ c_1^{-1}a_1 & 1 & 0 \\ -c_2^{-1}kc_1^{-1}a_1 + c_2^{-1}a_2 & 0 & 1 \end{pmatrix}.$$

(v) Check the preceding components using the alternative notation:

$$\underset{4}{\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right]} = \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right],$$

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right]_4 = \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right],$$

$$\underset{1}{\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right]} = \left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right],$$

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}\right]_1 = \left[\begin{pmatrix} c_1 & 0 \\ k & c_2 \end{pmatrix}^{-1}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right].$$

19. (i) Show that the location group G_1 (Problem 2) is a normal subgroup of the positive affine group G_7 (Problem 11).

(ii) Show that the factor group G_7/G_1 can be represented by the positive linear group G_6 (Problem 10).

(iii) Show that

$$\underset{6}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{11} & c_{12} \\ 0 & c_{21} & c_{22} \end{pmatrix},$$

$$\underset{6}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{11} & c_{12} \\ 0 & c_{21} & c_{22} \end{pmatrix},$$

a consequence of the normality of H_1.

(iv) Show that

$$\underset{1}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ a_1 & 1 & 0 \\ a_2 & 0 & 1 \end{pmatrix},$$

$$\underset{1}{[g]} = \begin{pmatrix} 1 & 0 & 0 \\ c^{11}a_1 + c^{12}a_2 & 1 & 0 \\ c^{21}a_1 + c^{22}a_2 & 0 & 1 \end{pmatrix},$$

where

$$\begin{pmatrix} c^{11} & c^{12} \\ c^{21} & c^{22} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1}$$

is an inverse matrix.

(v) Check the decompositions using the alternative notation:

$$\left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}\right] = \left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right]\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}\right]$$

$$= \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}\right]\left[\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right].$$

***20.** If H_1 is normal in G and $H_1 \subset H_2 \subset G$, where H_2 is a subgroup of G, show that H_1 is normal in H_2. Compare Problems 18 and 19 and check $G_1 \subset G_5 \subset G_7$.

***21.** (i) For the measurement model with normal error determine the conditional error probability distribution and the conditional structural distribution given the information $\mu = \mu_0$ (see Figure 9).

(ii) The structural model constructed with the information $\mu = \mu_0$ is

$$\Pi f(e_i) \, \Pi \, de_i,$$

$$y_1 = x_1 - \mu_0 = \sigma e_1$$

$$\vdots$$

$$y_n = x_n - \mu_0 = \sigma e_n,$$

where the error distribution is that of a standard normal sample. This is a simple measurement model with multiplicative error (Problem 19 in Chapter One). Determine the error probability distribution and the structural distribution for σ.

(iii) Compare the distributions for σ.

*22. In a rare application of a structural model the value of the quantity θ may have occurred as a realized value from a random process with known distribution $p(\theta)\,d\nu(\theta)$. Primary interest would center on the information concerning θ from the structural model itself. In certain circumstances, however, there might also be interest in the information from the *combined* processes. The composite model is

$$p(\theta)\,d\nu(\theta),$$
$$f(E)\,dm(E),$$
$$X = \theta E.$$

The model has a distribution describing the process that produces the θ value in the structural equation; it has a distribution describing the error process that produced the E value in the structural equation; and it has a structural equation linking the known X and the unknowns θ and E.

(i) Consider the contours corresponding to the information that specifies the value of X. Show that this information is based on the partition of $G \times \mathfrak{X}$ into *orbits:*

$$G(\theta, E) = \{(\theta g^{-1}, gE):\ g \in G\}.$$

(ii) Show that $[E]$ is a transformation variable for these orbits; show that $([X], D(X))$ is the reference point; and show that X indexes the orbits.

(iii) Show that

$$d\nu(\theta)\,dm(E)$$

is an invariant differential.

(iv) Show that the conditional distribution given available information is

$$k^*(X)\tilde{f}([E]D)p([X][E]^{-1})\,d\mu([E])$$

in terms of $[E]$ or is

$$k^*(X)\tilde{f}(\theta^{-1}X)p(\theta)\,\Delta([X])\,d\nu(\theta)$$

in terms of θ; $k^*(X)$ is the normalizing constant.

*23. Let $f_1(E)\,dE$, $X = \theta E$ and $f_2(F)\,dF$, $Y = \theta F$ be two structural models with a common quantity θ in a group G.

(i) Check that the composite model,

$$f_1(E)f_2(F)\,dE\,dF,$$
$$(X, Y) = \theta(E, F),$$

is a structural model.

(ii) Show that the structural distribution for θ from the composite model,

$$\tilde{g}^*(\theta : X, Y)\,d\nu(\theta),$$

can be obtained from the joint distribution of $\theta = \theta_1$ from the first model and $\theta = \theta_2$ from the second model by imposing the condition $\theta_1 = \theta_2$ relative to the right invariant differential.

CHAPTER THREE

Linear Models

The measurement model was developed to describe a system with all controllable variables held constant: the response variable was real valued; the internal error as it expressed itself in the response was distributed with known form.

In this chapter two structural models are developed as different extensions of the measurement model. The *regression model* handles a broad class of systems in which the controllable variables are allowed to vary or are manipulated. The *progression model* provides an extension in a different direction and handles a rather special kind of system with vector-valued response variable, special in being *progressively* structured in terms of error components. The range of applications of the second model is limited, but it supplies some of the notation and method to be used for a more comprehensive model that is developed in later chapters.

The regression model and the progression model can be combined in a single general composite model. A succession of problems presents this extension.

THE REGRESSION MODEL

1 EXAMPLES

Consider a stable system having a real-valued response. Suppose that selected controllable variables are subject to manipulation and that the response component of internal error has a known distribution $f(e)\,de$ on R^1. Suppose also that twelve performances of the system have been made and $y_1, \ldots, y_{12}$ are the observed values of the response variable.†

1.1 If the controllable variables do not affect the response level, then the measurement model is applicable. Let μ designate the general response level

† In the presence of controllable variables a response variable is typically designated by y and a controllable variable by x.

and σ designate the response scaling of error. The structural equation can be written in the special notation of Chapter One or in the matrix form indicated by Problem 26 in Chapter One; the matrix form is more convenient for generalization here. The structural equation is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} & y_{12} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 \\ \mu & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{bmatrix}.$$

The positive affine transformations

$$G = \left\{ \begin{bmatrix} 1 & 0 \\ a & c \end{bmatrix} : \begin{matrix} -\infty < a < \infty \\ 0 < c < \infty \end{matrix} \right\}$$

form a group under matrix multiplication. The orbit of a point $\mathbf{y} = (y_1, \ldots, y_{12})'$ in R^{12} is the half-plane containing $\mathbf{y}$ and subtended by the one-dimensional subspace $L(\mathbf{1}) = \{a\mathbf{1}: \ -\infty < a < \infty\}$. The orbit is contained in the two-dimensional subspace

$$L(\mathbf{1}, \mathbf{y}) = \left\{ a_1\mathbf{1} + a_2\mathbf{y}: \begin{matrix} -\infty < a_1 < \infty \\ -\infty < a_2 < \infty \end{matrix} \right\}$$

but consists of points with positive coefficient for $\mathbf{y}$:

$$L^+(\mathbf{1}; \mathbf{y}) = \left\{ a\mathbf{1} + c\mathbf{y}: \begin{matrix} -\infty < a < \infty \\ 0 < c < \infty \end{matrix} \right\}.$$

(See Figure 1.)

1.2 Now suppose that nine of the performances were chosen at random and given a certain treatment, and that the remaining three were given no treatment; designate those with no treatment by 1, 2, 3 and those with treatment by 4, ..., 12. Let β_1 designate the response level with no treatment, and β_2 designate the *increase* in level from no-treatment to with-treatment. The structural equation can be expressed as

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} & y_{12} \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_1 & \beta_2 & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{bmatrix}.$$

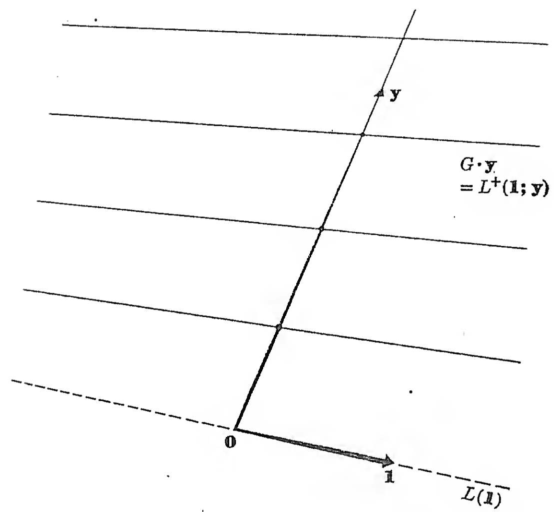


Figure 1 The subtending one-dimensional subspace $L(\mathbf{1})$; The orbit $G \cdot \mathbf{y} = L^+(\mathbf{1}; \mathbf{y})$, a positive half of the two-dimensional subspace $L(\mathbf{1}, \mathbf{y})$.

An additional row has been adjoined to the error and response vectors to permit the continued use of matrix multiplication.

The transformations

$$G = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_1 & a_2 & c \end{bmatrix} : \begin{matrix} -\infty < a_u < \infty \\ 0 < \ c < \infty \end{matrix} \right\}$$

form a group under matrix multiplication:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ A_1 & A_2 & C \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_1 & a_2 & c \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ A_1 + Ca_1 & A_2 + Ca_2 & Cc \end{bmatrix},$$

$$i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_1 & a_2 & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -c^{-1}a_1 & -c^{-1}a_2 & c^{-1} \end{bmatrix}.$$

This is a simple generalization of the positive affine group; it is a *regression-scale group*. The orbit of a point $\mathbf{y} = (y_1, \ldots, y_{12})'$ is the set

$$G \cdot \mathbf{y} = \left\{ a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + c\mathbf{y}: \quad \begin{matrix} -\infty < a_u < \infty \\ 0 < c < \infty \end{matrix} \right\},$$

where†

$$\mathbf{v}_1 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)',$$
$$\mathbf{v}_2 = (0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1)'.$$

(See Figure 2.) The orbit is a *half three-space*; the orbit is contained in the *three-dimensional subspace*

$$L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{y}) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{y}: \quad -\infty < a_u < \infty\}$$

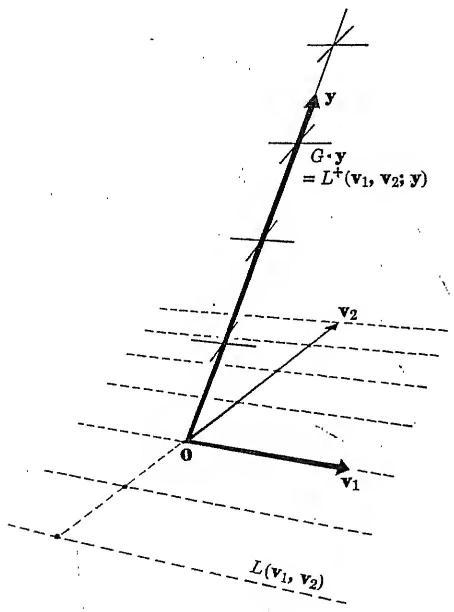


Figure 2 The subtending two-dimensional subspace $L(\mathbf{v}_1, \mathbf{v}_2)$. The orbit $G \cdot \mathbf{y} = L^+(v_1, v_2; \mathbf{y})$, a positive half of the three-dimensional subspace $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{y})$.

† The elements of G are transformations on R^{12}; the dot designates a transformation acting on $\mathbf{y}$ in R^{12}. A transformation as represented by a matrix can act by matrix multiplication provided the additional vectors are adjoined to $\mathbf{y}$.

but consists of the *half* with positive coefficient for $\mathbf{y}$:

$$L^+(\mathbf{v}_1, \mathbf{v}_2; \mathbf{y}) = \left\{ a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + c\mathbf{y}: \quad \begin{matrix} -\infty < a_u < \infty \\ 0 < c < \infty \end{matrix} \right\}.$$

The orbit is subtended by the two-dimensional subspace

$$L(\mathbf{v}_1, \mathbf{v}_2) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2\}$$

and consists of the *positive* translates

$$L(\mathbf{v}_1, \mathbf{v}_2) + c\mathbf{y}$$

of that two-dimensional subspace. Note that points $\mathbf{y}$ in the subspace $L(\mathbf{v}_1, \mathbf{v}_2)$ are implicitly excluded; compare the measurement model, Section 12, Chapter One.

The characteristics of the process can be described in various ways. As alternative quantities describing the process, let α_1 designate the *average* response level corresponding to the twelve performances, and α_2 designate the increase from the average level for no treatment to the average level with treatment. The structural equation can then be expressed as

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} & y_{12} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_1 & \alpha_2 & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{bmatrix}.$$

The general level corresponding to the first three performances is

$$\alpha_1 - \tfrac{3}{4}\alpha_2 = \beta_1,$$

and corresponding to the remaining nine is

$$\alpha_1 + \tfrac{1}{4}\alpha_2 = \beta_1 + \beta_2.$$

The orbit of a point $\mathbf{y}$ is of course the same as before. The orbit is

$$G \cdot \mathbf{y} = \left\{ a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + c\mathbf{y}: \quad \begin{matrix} -\infty < a_u < \infty \\ 0 < c < \infty \end{matrix} \right\}$$
$$= L^+(\mathbf{w}_1, \mathbf{w}_2; \mathbf{y}\},$$

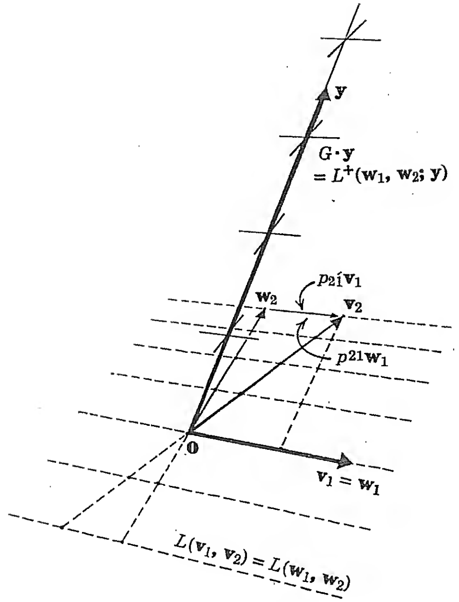


Figure 3 The subtending two-dimensional subspace $L(w_1, w_2) = L(v_1, v_2)$. The orthogonalized vector w_2 is obtained from v_2 by subtracting the component $p_{21}v_1 = p^{21}w_1$. The orbit $G \cdot y = L^{+}(w_1, w_2; y)$, a positive half of the three-dimensional subspace $L(w_1, w_2, y)$.

where

$$w_1 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)',$$

$$w_2 = (-\tfrac{3}{4}, -\tfrac{3}{4}, -\tfrac{3}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4})'.$$

The *basis* v_1, v_2 for the subtending space has been replaced by the new basis w_1, w_2 (see Figure 3). Note that w_2 is orthogonal to $w_1 (= v_1)$.

The new matrix expression for the observation vector y can be expressed in terms of the old:

$$\begin{pmatrix} w_1' \\ w_2' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \\ y' \end{pmatrix};$$

in brief, the change of basis for the subtending subspace is

$$\begin{pmatrix} w_1' \\ w_2' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\tfrac{3}{4} & 1 \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \end{pmatrix}$$

$$= P \begin{pmatrix} v_1' \\ v_2' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -p_{21} & 1 \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \end{pmatrix}.$$

The new structural equation can be related to the old:

$$\begin{pmatrix} w_1' \\ w_2' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \\ y' \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_1 & \beta_2 & \sigma \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \\ e' \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_1 & \beta_2 & \sigma \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$

$$\cdot \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1' \\ v_2' \\ e' \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_1 & \alpha_2 & \sigma \end{pmatrix} \cdot \begin{pmatrix} w_1' \\ w_2' \\ e' \end{pmatrix}.$$

And the new *quantity* in terms of the old can then be extracted:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_1 & \alpha_2 & \sigma \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_1 & \beta_2 & \sigma \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\tfrac{3}{4} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1};$$

or

$$(\alpha_1, \alpha_2) = (\beta_1, \beta_2) \begin{pmatrix} 1 & 0 \\ -\tfrac{3}{4} & 1 \end{pmatrix}^{-1} = (\beta_1, \beta_2) \begin{pmatrix} 1 & 0 \\ \tfrac{3}{4} & 1 \end{pmatrix}.$$

$$= (\beta_1, \beta_2) P^{-1} = (\beta_1, \beta_2) \begin{pmatrix} 1 & 0 \\ p^{21} & 1 \end{pmatrix}.$$

Note that $\alpha_2 = \beta_2$. The alternative structural equation with its orthogonal basis vectors $\mathbf{w}_1, \mathbf{w}_2$ has some advantages for later analysis.

1.3 Now suppose that three of the nine treatment performances were chosen at random and given a level 0 of a variable x integral to the treatment, that three of the remaining six performances were chosen at random and given the level $x = 1$, and the remaining three were given the level $x = 2$. Let the performances be numbered 4, 5, 6 for level $x = 0$; 7, 8, 9 for $x = 1$; and 10, 11, 12 for $x = 2$.

If the additional variable x affects the response level linearly with coefficient β_3, then the structural equation can be expressed as

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} & y_{12} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta_1 & \beta_2 & \beta_3 & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{bmatrix}.$$

The transformations

$$G = \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a_1 & a_2 & a_3 & c \end{bmatrix} : \begin{matrix} -\infty < a_u < \infty \\ \\ 0 < c < \infty \end{matrix} \right\}$$

form a group under matrix multiplication, an example of a regression-scale group. The orbit of a point $\mathbf{y} = (y_1, \ldots, y_{12})'$ is the set

$$G \cdot \mathbf{y} = \left\{ a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + c\mathbf{y} : \begin{matrix} -\infty < a_u < \infty \\ 0 < c < \infty \end{matrix} \right\}$$

$$= L^+(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3; \mathbf{y}),$$

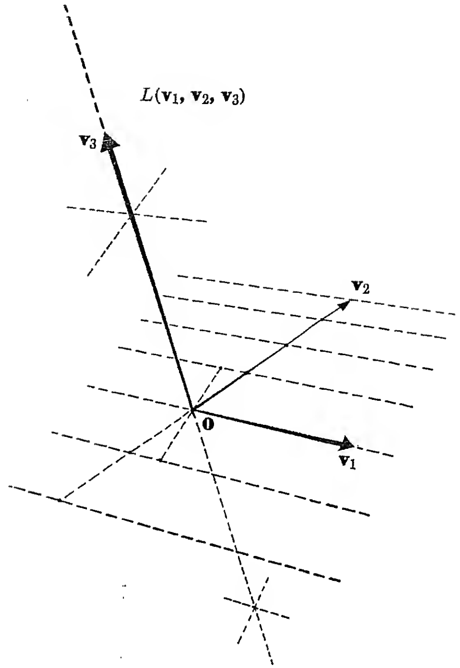


Figure 4 The three-dimensional subspace $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ which subtends the half four-space $G \cdot \mathbf{y} = L^+(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3; \mathbf{y})$.

where

$$\mathbf{v}_3 = (0, 0, 0, 0, 0, 0, 1, 1, 1, 2, 2, 2)'.$$

(See Figure 4.) The orbit is a *half four-space*; it is contained in the *four-dimensional subspace* $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{y})$ but consists of the *half* with positive coefficient for $\mathbf{y}$. The orbit is subtended by the three-dimensional subspace

$$L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3\},$$

and it consists of the *positive* translates

$$L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) + c\mathbf{y}$$

of that three-dimensional subspace.

As alternative quantities describing the process, let α_1 designate the *average* response level corresponding to the twelve performances, let α_2 designate the increase from the *average* level for no-treatment to the *average*

level with-treatment, and let α_3 be the coefficient for *change* of level with respect to the variable x. The structural equation can then be expressed as

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & y_{11} & y_{12} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \alpha_1 & \alpha_2 & \alpha_3 & \sigma \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{bmatrix}.$$

The orbit of a point $\mathbf{y}$ is of course the same as before. The orbit is

$$G \cdot \mathbf{y} = \left\{ a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + a_3\mathbf{w}_3 + c\mathbf{y} : \begin{array}{r} -\infty < a_u < \infty \\ 0 < c < \infty \end{array} \right\}$$

$$= L^+(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3; \mathbf{y}),$$

where

$$\mathbf{w}_3 = (0, 0, 0, -1, -1, -1, 0, 0, 0, 1, 1, 1)'.$$

The basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ for the subtending subspace has been replaced by the new basis $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ (see Figure 5). Note that $\mathbf{w}_3$, $\mathbf{w}_2$ and $\mathbf{w}_1 (= \mathbf{v}_1)$ are mutually orthogonal.

The new matrix expression containing the observation vector $\mathbf{y}$ can be expressed in terms of the old:

$$\begin{bmatrix} \mathbf{w}_1' \\ \mathbf{w}_2' \\ \mathbf{w}_3' \\ \mathbf{y}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \mathbf{v}_3' \\ \mathbf{y}' \end{bmatrix};$$

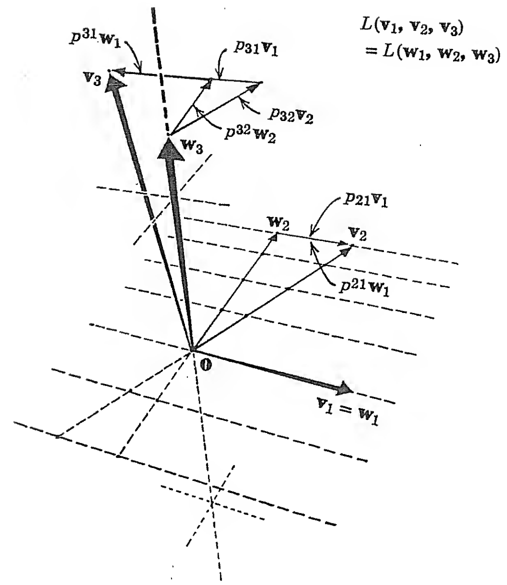


Figure 5 The subspace $L(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$ which subtends the orbit $G \cdot \mathbf{y} = L^+(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3; \mathbf{y})$. The orthogonalized vector $\mathbf{w}_3$ is obtained from $\mathbf{v}_3$ by subtracting components $p_{31}\mathbf{v}_1$, $p_{32}\mathbf{v}_2$ or equivalently by subtracting components $p^{31}\mathbf{w}_1$, $p^{32}\mathbf{w}_2$ (note $p_{32} = p^{32}$). In the example $\mathbf{w}_3 = \mathbf{v}_3 - \mathbf{v}_2$; the diagram illustrates the more general case in which $\mathbf{w}_3$ is formed from $\mathbf{v}_3$ by removing both $\mathbf{v}_1$- and $\mathbf{v}_2$-components.

in brief, the change of basis for the subtending subspace is

$$\begin{bmatrix} \mathbf{w}_1' \\ \mathbf{w}_2' \\ \mathbf{w}_3' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \mathbf{v}_3' \end{bmatrix}$$

$$= P \begin{bmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \mathbf{v}_3' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -p_{21} & 1 & 0 \\ -p_{31} & -p_{32} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \mathbf{v}_3' \end{bmatrix}.$$

The new structural equation can be related to the old equation and the new

quantity in terms of the old quantity can then be extracted:

$$
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \alpha_1 & \alpha_2 & \alpha_3 & \sigma \end{bmatrix}
= \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta_1 & \beta_2 & \beta_3 & \sigma \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1},
$$

or, in brief,

$$
(\alpha_1, \alpha_2, \alpha_3) = (\beta_1, \beta_2, \beta_3) \begin{bmatrix} 1 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}^{-1} = (\beta_1, \beta_2, \beta_3) \begin{bmatrix} 1 & 0 & 0 \\ \frac{3}{4} & 1 & 0 \\ \frac{3}{4} & 1 & 1 \end{bmatrix}
$$

$$
= (\beta_1, \beta_2, \beta_3)P^{-1} = (\beta_1, \beta_2, \beta_3) \begin{bmatrix} 1 & 0 & 0 \\ p^{21} & 1 & 0 \\ p^{31} & p^{32} & 1 \end{bmatrix}.
$$

Note that $\alpha_3 = \beta_3$. The alternative structural equation with its orthogonal basis vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ is convenient for some analysis in later sections.

2 THE MODEL

Consider again a stable system with a real-valued response y. Suppose that selected controllable variables are subject to manipulation and that the response component of internal error has a known distribution $f(e)\,de$ on R^1.

Also suppose that there have been n performances of the system and the observations on the response variable are $\mathbf{y} = (y_1, \ldots, y_n)'$. And suppose that various controllable variables have been manipulated and that information concerning the system presents the sequence of general response levels as linear in *structural vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_r$: The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ record values of treatment indicator-variables, or values of controllable variables, or values of combinations of these variables; compare with the examples in Section 1. As quantities characteristic of the system let σ be the response scaling of error, and let $\beta_1, \ldots, \beta_r$ be the coefficients that present the response levels in terms of the structural vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$. The system and the n performances can then be described by the

Regression Model

$$
\prod_1^n f(e_i) \prod_1^n de_i,
$$

$$
\begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{y}' \end{bmatrix}
= \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & & \cdot \\ 0 & & 1 & 0 \\ \beta_1 & \cdots & \beta_r & \sigma \end{bmatrix}
\begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{e}' \end{bmatrix}.
$$

The model has two parts: an error distribution which describes the internal operation of the system (with $\mathbf{e}$ as a *variable*); and a structural equation in which a realized vector $\mathbf{e}$ from the error distribution has determined the relation between the response observations $\mathbf{y}$ and the unknown values $\beta_1, \ldots, \beta_r, \sigma$ for the system characteristics (with $\mathbf{e}$ as a *constant*).

The notation can be made more compact by letting

$$
Y = \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \vdots \\ \mathbf{v}_r' \\ \mathbf{y}' \end{bmatrix} = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \\ y_1 & \cdots & y_n \end{bmatrix}
$$

designate the *response vector* $\mathbf{y}$ with appended structural vectors $\mathbf{v}_1, \ldots \mathbf{v}_r$; by letting

$$
E = \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{e}' \end{bmatrix} = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \\ e_1 & \cdots & e_n \end{bmatrix}
$$

designate the *error vector* $\mathbf{e}$ with appended structural vectors; by letting

$$
\theta = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & & \cdot \\ 0 & & 1 & 0 \\ \beta_1 & \cdots & \beta_r & \sigma \end{bmatrix}
$$

designate the composite *quantity* in matrix array; and by letting

$$f(E)\,dE = \prod f(e_i) \prod de_i$$

designate the *error distribution*. The regression model can then be written

$$f(E)\,de,$$

$$Y = \theta E.$$

The transformation θ is an element of the *regression-scale group*

$$G = \left\{ g = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ a_1 & \cdots & a_r & c \end{bmatrix} : \begin{array}{c} -\infty < a_u < \infty \\ 0 < c < \infty \end{array} \right\},$$

with group properties

$$\begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ A_1 & \cdots & A_r & C \end{bmatrix} \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ a_1 & \cdots & a_r & c \end{bmatrix} = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ A_1 + Ca_1 & \cdots & A_r + Ca_r & Cc \end{bmatrix},$$

$$i = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ a_1 & \cdots & a_r & c \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ -c^{-1}a_1 & \cdots & -c^{-1}a_r & c^{-1} \end{bmatrix}.$$

The orbit of a point Y is the set

$$GY = \{gY: \quad g \in G\},$$

or equivalently the set

$$G \cdot \mathbf{y} = \left\{ a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r + c\mathbf{y}: \quad \begin{array}{c} -\infty < a_u < \infty \\ 0 < c < \infty \end{array} \right\}$$

$$= L^+(\mathbf{v}_1, \ldots, \mathbf{v}_r; \mathbf{y}).$$

Suppose now that $n \geq r + 1$ *and that* $\mathbf{v}_1, \ldots \mathbf{v}_r$ *are linearly independent*; this avoids trivial cases with more quantities than effective measurements. The orbit is a *half* $(r + 1)$*-space*; it is contained in the $(r + 1)$*-dimensional subspace*

$$L(\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{y}) = \{a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r + a_{r+1}\mathbf{y}: \quad -\infty < a_u < \infty\}$$

but consists of the positive *half* corresponding to positive coefficient for $\mathbf{y}$. The orbit is subtended by the r-dimensional subspace

$$L(\mathbf{v}_1, \ldots, \mathbf{v}_r),$$

and it consists of the positive translates

$$L(\mathbf{v}_1, \ldots, \mathbf{v}_r) + c\mathbf{y}$$

of that subspace; see Figures 1 and 2. *The points* $\mathbf{y}$ *in the subspace* $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ *are implicitly excluded without loss of essential generality in the sequel.*

A point $\mathbf{y}$ is carried into a point

$$\tilde{\mathbf{y}} = a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r + c\mathbf{y}$$

by a transformation g. The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{y}$ are linearly independent; there is then no alternative choice for a transformation carrying $\mathbf{y}$ into $\tilde{\mathbf{y}}$. It follows that G is unitary on R^n (subtending subspace excluded). And it follows then that *the regression model is a structural model.*

3 A TRANSFORMATION VARIABLE

Consider the choice of a transformation variable $[Y]$ to describe the position of a point Y on its orbit. For the measurement model the *location* variable $\bar{x}$ gave the projection $\bar{x}\mathbf{1}$ of the vector $\mathbf{x}$ onto the one-dimensional subspace $L(\mathbf{1})$, the *scale* variable s_x gave the distance of $\mathbf{x}$ from $L(\mathbf{1})$ (units of length $(n - 1)^{1/2}$), and the *transformation* $[\bar{x}, s_x]$ gave position (see Figure 6). A transformation variable can be defined for the linear regression model in an analogous way: *Location* can be described by the projection of $\mathbf{y}$ into the subtending subspace $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$, *scale* can be given by the distance of $\mathbf{y}$ from the subspace, and *position* by combining these into a transformation matrix.

Consider a point $\mathbf{y}$ (or Y) in R^n, and let

$$b_1(\mathbf{y})\mathbf{v}_1 + \cdots + b_r(\mathbf{y})\mathbf{v}_r$$

be the *projection* of $\mathbf{y}$ into the r-dimensional subspace $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$: *The projection of* $\mathbf{y}$ *into the subspace* $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ *is that vector*

$$b_1\mathbf{v}_1 + \cdots + b_r\mathbf{v}_r$$

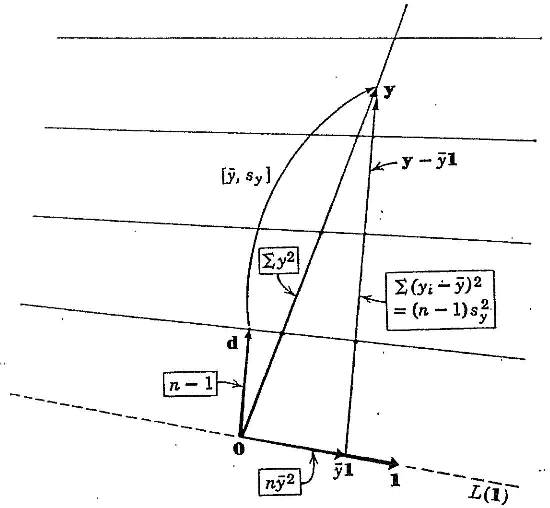


Figure 6 The vector $\mathbf{y}$, and its projection $\bar{y}\mathbf{1}$ on the one-dimensional subspace $L(\mathbf{1})$. Squared lengths of vectors are recorded.

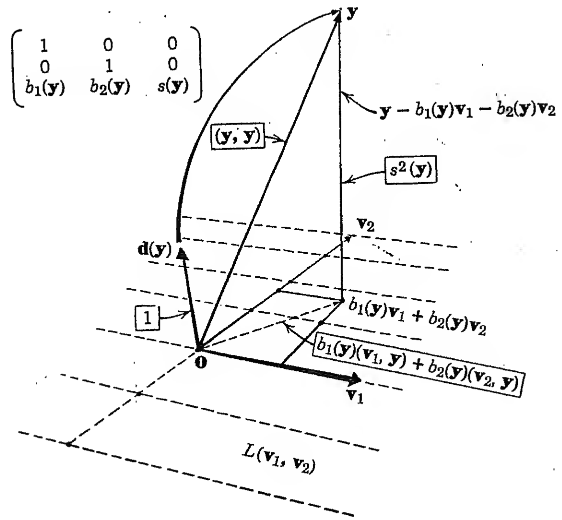


Figure 7 The projection $b_1(\mathbf{y})\mathbf{v}_1 + b_2(\mathbf{y})\mathbf{v}_2$ of $\mathbf{y}$ into the two-dimensional subspace $L(\mathbf{v}_1, \mathbf{v}_2)$. Squared lengths of vectors are recorded.

in the subspace for which the residual vector

$$\mathbf{y} - (b_1\mathbf{v}_1 + \cdots + b_r\mathbf{v}_r)$$

is orthogonal to each of $\mathbf{v}_1, \ldots, \mathbf{v}_r$, *and hence to each vector in* $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$. (See Figure 7.) The *orthogonality conditions* in the definition are

$$\begin{gathered} (\mathbf{y} - b_1\mathbf{v}_1 - \cdots - b_r\mathbf{v}_r, \mathbf{v}_1) = 0 \\ \cdot \\ \cdot \\ \cdot \\ (\mathbf{y} - b_1\mathbf{v}_1 - \cdots - b_r\mathbf{v}_r, \mathbf{v}_r) = 0, \end{gathered}$$

where $(\mathbf{x}, \mathbf{y})$ designates the *inner product*

$$(\mathbf{x}, \mathbf{y}) = x_1y_1 + \cdots + x_ny_n = (\mathbf{y}, \mathbf{x})$$

of the vectors $\mathbf{x}$ and $\mathbf{y}$. The inner product is linear in each argument:

$$\begin{aligned} (a_1\mathbf{x}_1 + a_2\mathbf{x}_2, \mathbf{y}) &= a_1(\mathbf{x}_1, \mathbf{y}) + a_2(\mathbf{x}_2, \mathbf{y}), \\ (\mathbf{x}, b_1\mathbf{y}_1 + b_2\mathbf{y}_2) &= b_1(\mathbf{x}, \mathbf{y}_1) + b_2(\mathbf{x}, \mathbf{y}_2); \end{aligned}$$

accordingly, the conditions can be rearranged to give the *orthogonality equations*:

$$\begin{gathered} (\mathbf{v}_1, \mathbf{v}_1)b_1 + \cdots + (\mathbf{v}_1, \mathbf{v}_r)b_r = (\mathbf{v}_1, \mathbf{y}) \\ \cdot \\ \cdot \\ \cdot \\ (\mathbf{v}_r, \mathbf{v}_1)b_1 + \cdots + (\mathbf{v}_r, \mathbf{v}_r)b_r = (\mathbf{v}_r, \mathbf{y}). \end{gathered}$$

The inner products as they appear in the preceding array form the first r rows of the matrix product:

$$YY' = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \\ y_1 & \cdots & y_n \end{bmatrix} \begin{bmatrix} v_{11} & \cdots & v_{r1} & y_1 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ v_{1n} & \cdots & v_{rn} & y_n \end{bmatrix}$$

$$= \left[\begin{array}{cccc} (\mathbf{v}_1, \mathbf{v}_1) & \cdots & (\mathbf{v}_1, \mathbf{v}_r) & (\mathbf{v}_1, \mathbf{y}) \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ (\mathbf{v}_r, \mathbf{v}_1) & \cdots & (\mathbf{v}_r, \mathbf{v}_r) & (\mathbf{v}_r, \mathbf{y}) \\ \hdashline (\mathbf{y}, \mathbf{v}_1) & \cdots & (\mathbf{y}, \mathbf{v}_r) & (\mathbf{y}, \mathbf{y}) \end{array}\right].$$

Let the first r rows of Y be designated

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \end{bmatrix};$$

the orthogonality equations can then be written:

$$VV' \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_r \end{bmatrix} = V\mathbf{y}.$$

The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ have been assumed linearly independent: the matrix V is of rank r; and the matrix VV' is then of rank r, hence is nonsingular. The orthogonality equations can then be solved *uniquely* to give

$$\begin{bmatrix} b_1(\mathbf{y}) \\ \cdot \\ \cdot \\ \cdot \\ b_r(\mathbf{y}) \end{bmatrix} = (VV')^{-1}V\mathbf{y}.$$

The projection of $\mathbf{y}$ *into the subspace* $L(\mathbf{v}_1, \ldots \mathbf{v}_r)$ *is*

$$b_1(\mathbf{y})\mathbf{v}_1 + \cdots + b_r(\mathbf{y})\mathbf{v}_r,$$

where the coefficients $b_1(\mathbf{y}), \ldots, b_r(\mathbf{y})$ *are given by*

$$\begin{bmatrix} b_1(\mathbf{y}) \\ \cdot \\ \cdot \\ \cdot \\ b_r(\mathbf{y}) \end{bmatrix} = (VV')^{-1}V\mathbf{y};$$

the coefficients are called *the regression coefficients of* $\mathbf{y}$ *on* $\mathbf{v}_1, \ldots, \mathbf{v}_r$. In the case of a single $\mathbf{v}$ the regression coefficient is

$$b(\mathbf{y}) = \frac{(\mathbf{v}, \mathbf{y})}{(\mathbf{v}, \mathbf{v})};$$

the general form is analogous: *the inner products with* $\mathbf{y}$ *multiplied on the left by the inverse of the inner-product matrix.*

The projection of $\mathbf{y}$ into the subspace $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ can be defined alternatively as *that point*

$$b_1\mathbf{v}_1 + \cdots + b_r\mathbf{v}_r$$

in the subspace at minimum distance from $\mathbf{y}$. Let $b_1(\mathbf{y}), \ldots, b_r(\mathbf{y})$ be the coefficients obtained by solving the orthogonality equations. The difference vector

$$\mathbf{y} - \sum_1^r b_u\mathbf{v}_u$$

can be represented as a sum

$$\left(\mathbf{y} - \sum_1^r b_u(\mathbf{y})\mathbf{v}_u\right) + \sum_1^r (b_u(\mathbf{y}) - b_u)\mathbf{v}_u$$

of two vectors that are orthogonal; accordingly, the squared length of the original vector,

$$\left|\mathbf{y} - \sum_1^r b_u\mathbf{v}_u\right|^2,$$

is equal† to the squared length of the first vector plus the squared length of the second vector:

$$\left|\mathbf{y} - \sum_1^r b_u(\mathbf{y})\mathbf{v}_u\right|^2 + \left|\sum_1^r (b_u(\mathbf{y}) - b_u)\mathbf{v}_u\right|^2.$$

Choosing $b_1, \ldots, b_r$ to minimize the length of the original vector is equivalent to choosing $b_1, \ldots, b_r$ to minimize the length of the second vector, but the second vector can be made equal to the *zero* vector by choosing $b_u = b_u(\mathbf{y})$. Thus the *projection* into the subspace is the *closest point* in the subspace (see Figure 8).

The residual vector is

$$\mathbf{y} - b_1(\mathbf{y})\mathbf{v}_1 - \cdots - b_r(\mathbf{y})\mathbf{v}_r.$$

Let $s(\mathbf{y})$ be the *residual length*:

$$\begin{aligned} s(\mathbf{y}) &= |\mathbf{y} - \textstyle\sum b_u(\mathbf{y})\mathbf{v}_u|, \\ s^2(\mathbf{y}) &= |\mathbf{y} - \textstyle\sum b_u(\mathbf{y})\mathbf{v}_u|^2 \\ &= (\mathbf{y} - \textstyle\sum b_u(\mathbf{y})\mathbf{v}_u, \mathbf{y} - \textstyle\sum b_u(\mathbf{y})\mathbf{v}_u); \end{aligned}$$

† Pythagoras. If $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, $(\mathbf{x}, \mathbf{y}) = 0$, then $|\mathbf{x} + \mathbf{y}|^2 = (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = (\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{y}) = (\mathbf{x}, \mathbf{x}) + (\mathbf{y}, \mathbf{y}) = |\mathbf{x}|^2 + |\mathbf{y}|^2$.

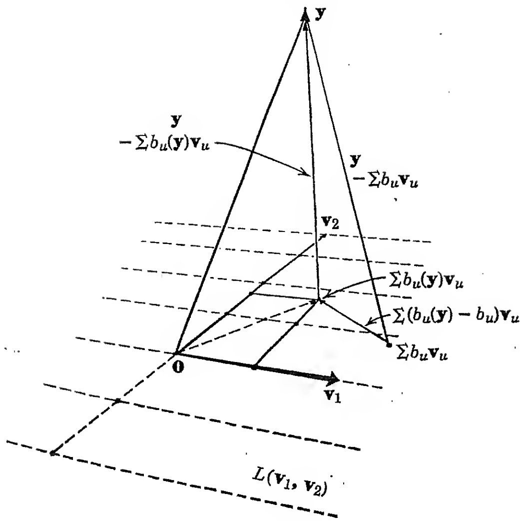


Figure 8 A general point $\Sigma\, b_u \mathbf{v}_u$ in the subspace $L(\mathbf{v}_1, \mathbf{v}_2)$, and the projection $\Sigma\, b_u(\mathbf{u})\mathbf{v}_u$ of $\mathbf{y}$ into the subspace $L(\mathbf{v}_1, \mathbf{v}_2)$. The projection point $\Sigma\, b_u(\mathbf{u})\mathbf{v}_u$ is the point in $L(\mathbf{v}_1, \mathbf{v}_2)$ that is closest to $\mathbf{y}$.

and let $\mathbf{d}(\mathbf{y})$ be the *unit residual vector*:

$$\mathbf{d}(\mathbf{y}) = s^{-1}(\mathbf{y})(\mathbf{y} - b_1(\mathbf{y})\mathbf{v}_1 - \cdots - b_r(\mathbf{y})\mathbf{v}_r).$$

The unit residual vector is orthogonal to the subspace $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$, has unit length and is a vector in $L^+(\mathbf{v}_1, \ldots, \mathbf{v}_r; \mathbf{y})$.

The vector $\mathbf{y}$ can be reconstructed from the regression coefficients and residual length:

$$\mathbf{y} = b_1(\mathbf{y})\mathbf{v}_1 + \cdots + b_r(\mathbf{y})\mathbf{v}_r + s(\mathbf{y})\,\mathbf{d}(\mathbf{y}).$$

This can be expressed in matrix notation:

$$Y = \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{y}' \end{bmatrix} = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ b_1(\mathbf{y}) & \cdots & b_r(\mathbf{y}) & s(\mathbf{y}) \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{d}'(\mathbf{y}) \end{bmatrix}$$

$$= [Y]\, D(Y).$$

The *variable*

$$[Y] = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & 0 \\ b_1(\mathbf{y}) & \cdots & b_r(\mathbf{y}) & s(\mathbf{y}) \end{bmatrix}$$

is an element of the regression-scale group; the matrix

$$D(Y) = \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{d}'(\mathbf{y}) \end{bmatrix} = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \\ d_1(\mathbf{y}) & \cdots & d_n(\mathbf{y}) \end{bmatrix}$$

then designates a vector on the orbit of $\mathbf{y}$, a vector $\mathbf{d}(\mathbf{y})$ which has *unit length* and is *orthogonal to the subtending subspace* $L(\mathbf{v}_1, \ldots, \mathbf{v}_r)$. It follows that $D(Y)$ is a *fixed* point, a reference point, on the orbit GY. And it follows then that $[Y]$ is a transformation variable. The transformation $[Y]$ as an element of a duplicate group G^* gives the *position* of Y on its orbit; see Figure 7.

The regression model can now be written

$$f(E)\, dE,$$
$$[Y] = \theta[E], \qquad D(Y) = D(E).$$

The structural equation conditional on the orbit has simple form:

$$b_1(\mathbf{y}) = \beta_1 + \sigma b_1(\mathbf{e})$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$b_r(\mathbf{y}) = \beta_r + \sigma b_r(\mathbf{e}),$$
$$s(\mathbf{y}) = \sigma s(\mathbf{e}).$$

The regression coefficients and residual length have produced a transformation variable with some convenient matrix properties. The regression coefficients and residual length are based on Euclidean distance:

$$|\mathbf{y} - \mathbf{x}| = [\sum (y_i - x_i)^2]^{1/2},$$

and on the related inner product:

$$(\mathbf{x}, \mathbf{y}) = \sum x_i y_i.$$

The use of other distance functions such as

$$d(\mathbf{y}, \mathbf{x}) = \sum |y_i - x_i|$$

can produce other transformation variables in an analogous manner. The distance function $\sum |y_i - x_i|$ has some computational advantages in that certain quadratic calculations for Euclidean distance are replaced by linear calculations; it is amenable to the linear programming algorithm and in some problems can be convenient for high-speed computation.

4 WITH ORTHOGONAL BASIS

In the examples in Section 1 structural vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ were successively introduced to describe successively more complex dependence of the general response level on controllable variables. Then, as an alternative, structural vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ were successively introduced to describe the same successively more complex dependence. The vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ were mutually orthogonal, and there was mention that this orthogonality had advantages in later analysis.

Now consider in general the regression model and suppose that the structural vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are in a natural order of decreasing intrusiveness, in the manner indicated by the examples. The corresponding sequence of orthogonal structural vectors can be constructed by the results in Section 3. For notation, consider the equation

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ w_{21} & \cdots & w_{2n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ w_{r1} & \cdots & w_{rn} \end{bmatrix} = PV$$

$$= \begin{bmatrix} 1 & & & & 0 \\ -p_{21} & 1 & & & \\ -p_{31} & -p_{32} & \cdot & & \\ \cdot & & \cdot & \cdot & \\ \cdot & & & \cdot & \cdot \\ -p_{r1} & \cdots & & -p_{r\,r-1} & 1 \end{bmatrix} \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ v_{21} & \cdots & v_{2n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \end{bmatrix}$$

Let p_{21} be the regression coefficient of $\mathbf{v}_2$ on $\mathbf{v}_1$; the vector $\mathbf{w}_2$ is then the corresponding residual vector and is orthogonal to $\mathbf{v}_1 = \mathbf{w}_1$. Let p_{31}, p_{32} be the regression coefficients of $\mathbf{v}_3$ on $\mathbf{v}_1, \mathbf{v}_2$; the vector $\mathbf{w}_3$ is then the corresponding residual vector and is orthogonal to $\mathbf{v}_1, \mathbf{v}_2$ and hence to $\mathbf{w}_1, \mathbf{w}_2$. Finally, let $p_{r1}, \ldots, p_{r\,r-1}$ be the regression coefficients of $\mathbf{v}_r$ on $\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}$; the vector $\mathbf{w}_r$ is then the corresponding residual vector and is orthogonal to $\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}$ and hence to $\mathbf{w}_1, \ldots, \mathbf{w}_{r-1}$. The matrix W records the orthogonal structural vectors $\mathbf{w}_1, \ldots, \mathbf{w}_r$ derived successively from the structural vectors in V.

The model can be presented in terms of the alternative basis for the subtending subspace. Let

$$\bar{Y} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ w_{r1} & \cdots & w_{rn} \\ y_1 & \cdots & y_n \end{bmatrix}$$

designate the response vector $\mathbf{y}$ but with orthogonal structural vectors appended; let

$$\bar{E} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ w_{r1} & \cdots & w_{rn} \\ e_1 & \cdots & e_n \end{bmatrix}$$

designate the error vector $\mathbf{e}$ but with orthogonal structural vectors appended; and let

$$\bar{\theta} = \begin{bmatrix} 1 & & 0 & 0 \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & & \cdot \\ 0 & & 1 & 0 \\ \alpha_1 & \cdots & \alpha_r & \sigma \end{bmatrix}$$

designate the composite quantity appropriate to the new orthogonal structural vectors. The regression model can then be expressed in the alternative form

$$f(\bar{E})\, d\bar{E},$$

$$\bar{Y} = \bar{\theta}\, \bar{E}.$$

The new matrix $\bar{Y}$ can be expressed in terms of the original matrix Y:

$$\bar{Y} = \left[\begin{array}{ccccc|c} 1 & & & & 0 & 0 \\ -p_{21} & 1 & & & & \\ -p_{31} & -p_{32} & 1 & & & \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ -p_{r1} & \cdots & -p_{r\,r-1} & & 1 & 0 \\ \hline 0 & \cdots & 0 & & 0 & 1 \end{array}\right] Y = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} Y,$$

And correspondingly $\bar{E}$ can be expressed in terms of E:

$$\bar{E} = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} E.$$

The new structural equation can be related to the old:

$$\begin{aligned} \bar{Y} = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} Y &= \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} \theta E \\ &= \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} \theta \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} E \\ &= \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} \theta \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix}^{-1} \cdot \bar{E}. \end{aligned}$$

The new quantity in terms of the old can then be extracted:

$$\bar{\theta} = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} \theta \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix}^{-1};$$

with σ common to $\bar{\theta}$ and θ, this can be expressed more briefly as

$$(\alpha_1, \ldots, \alpha_r) = (\beta_1, \ldots, \beta_r)P^{-1}.$$

The inverse of P has the following form:

$$P^{-1} = \begin{bmatrix} 1 & & & 0 \\ p^{21} & 1 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & 1 & \\ p^{r1} & & p^{r\,r-1} & 1 \end{bmatrix}.$$

Note then that $\alpha_r = \beta_r$. Compare with examples (1.2) and (1.3) in Section 1.

The general response level is of course the same regardless of the basis for the subtending subspace:

$$(\alpha_1, \ldots, \alpha_r)\begin{bmatrix} \mathbf{w}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{w}_r' \end{bmatrix} = (\beta_1, \ldots, \beta_r)P^{-1}\begin{bmatrix} \mathbf{w}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{w}_r' \end{bmatrix} = (\beta_1, \ldots, \beta_r)\begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \end{bmatrix}.$$

The general response level has *new quantities for a new basis.*

Consider the transformation variable in terms of the new matrix $\bar{Y}$:

$$[\bar{Y}] = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ a_1(\mathbf{y}) & \cdots & & a_r(\mathbf{y}) & s(\mathbf{y}) \end{bmatrix}.$$

The elements $a_1(\mathbf{y}), \ldots, a_r(\mathbf{y})$ designate the regression coefficients of $\mathbf{y}$ on $\mathbf{w}_1, \ldots, \mathbf{w}_r$:

$$\begin{bmatrix} a_1(\mathbf{y}) \\ \cdot \\ \cdot \\ \cdot \\ a_r(\mathbf{y}) \end{bmatrix} = \begin{bmatrix} (\mathbf{w}_1, \mathbf{w}_1) & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & & (\mathbf{w}_r, \mathbf{w}_r) \end{bmatrix}^{-1} \begin{bmatrix} (\mathbf{w}_1, \mathbf{y}) \\ \cdot \\ \cdot \\ \cdot \\ (\mathbf{w}_r, \mathbf{y}) \end{bmatrix},$$

$$a_u(\mathbf{y}) = \frac{(\mathbf{w}_u, \mathbf{y})}{(\mathbf{w}_u, \mathbf{w}_u)}.$$

The orthogonality of the structural vectors allows the regression coefficients to be calculated *individually*: a_u *is the regression coefficient of* $\mathbf{y}$ *on* $\mathbf{w}_u$. The projection of $\mathbf{y}$ into the subtending subspace by its definition depends only on $\mathbf{y}$ and on the subspace; it does not depend on the *basis* for that subspace. The residual vector

$$\begin{aligned} \mathbf{y} - a_1(\mathbf{y})\mathbf{w}_1 - \cdots - a_r(\mathbf{y})\mathbf{w}_r \\ = \mathbf{y} - b_1(\mathbf{y})\mathbf{v}_1 - \cdots - b_r(\mathbf{y})\mathbf{v}_r \\ = s(\mathbf{y})\,\mathbf{d}(\mathbf{y}) \end{aligned}$$

then has the same property; hence the residual length $s(\mathbf{y})$ and the unit residual vector $\mathbf{d}(\mathbf{y})$ also have this property. The matrix $\bar{Y}$ can now be written

$$\bar{Y} = [\bar{Y}]D(\bar{Y}) = [\bar{Y}]\begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ w_{r1} & \cdots & w_{rn} \\ d_1(\mathbf{y}) & \cdots & d_n(\mathbf{y}) \end{bmatrix};$$

(see Figure 9). Note that the reference point in matrix form $D(\bar{Y})$ records the unit residual vector but with the *orthogonal* structural vectors appended.

The new position $[\bar{Y}]$ in terms of the old can be extracted in the same manner as the new quantity $\bar{\theta}$ in terms of the old:

$$\bar{Y} = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} Y,$$

$$D(\bar{Y}) = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} D(Y),$$

$$\bar{Y} = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} Y = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} [Y]D(Y)$$

$$= \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} [Y] \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} D(Y)$$

$$= \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} [Y] \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix}^{-1} D(\bar{Y});$$

thus

$$[\bar{Y}] = \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix} [Y] \begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix}^{-1}.$$

With $s(\mathbf{y})$ common to $[\bar{Y}]$ and $[Y]$, this can be expressed more briefly as

$$(a_1(\mathbf{y}), \ldots, a_r(\mathbf{y})) = (b_1(\mathbf{y}), \ldots, b_r(\mathbf{y}))P^{-1}.$$

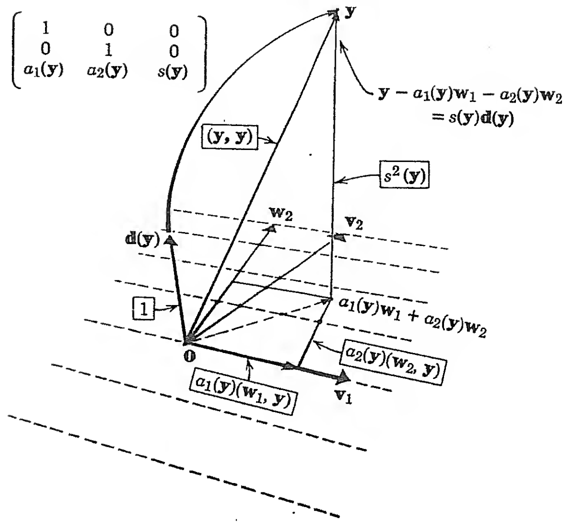


Figure 9 The projection $a_1(\mathbf{y})\mathbf{w}_1 + a_2(\mathbf{y})\mathbf{w}_2$ of $\mathbf{y}$ into the two-dimensional subspace $L(\mathbf{w}_1, \mathbf{w}_2) = L(\mathbf{v}_1, \mathbf{v}_2)$. Squared lengths of vectors are recorded.

These equations have the same form as the corresponding equations for $\bar{\theta}$ in terms of θ, and $(\alpha_1, \ldots, \alpha_r)$ in terms of $(\beta_1, \ldots, \beta_r)$. Note that $a_r(\mathbf{y}) = b_r(\mathbf{y})$.

The regression model in alternative form can now be expressed with composite structural equation:

$$f(\bar{E})\,d\bar{E},$$

$$[\bar{Y}] = \bar{\theta}[\bar{E}], \qquad D(\bar{Y}) = D(\bar{E}).$$

The structural equation conditional on the orbit can be written

$$a_1(\mathbf{y}) = \alpha_1 + \sigma a_1(\mathbf{e})$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$a_r(\mathbf{y}) = \alpha_r + \sigma a_r(\mathbf{e}),$$
$$s(\mathbf{y}) = \sigma s(\mathbf{e}).$$

The orthogonal structural vectors $\mathbf{w}_1, \ldots, \mathbf{w}_r$ can provide directions for the first r of a new set of axes. For the first axis, $a_1(\mathbf{y})$ measures distance in

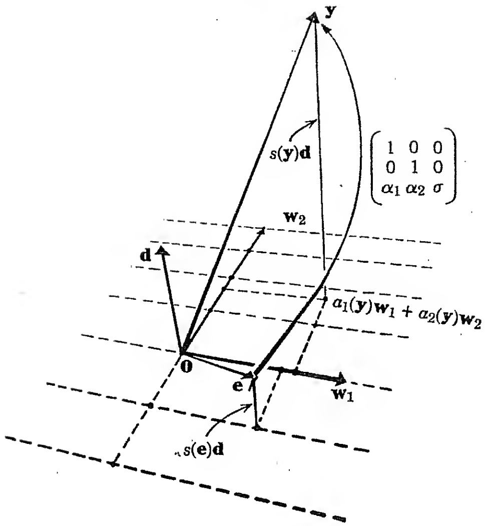


Figure 10 The observation $\bar{Y}$ as y; the error $\bar{E}$ as e; the transformation

$$\bar{\theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_1 & \alpha_2 & \sigma \end{bmatrix}.$$

units of length $|\mathbf{w}_1|$; the coordinate for the first axis is then

$$a_1(\mathbf{y})\,|\mathbf{w}_1| = \frac{(\mathbf{w}_1, \mathbf{y})}{(\mathbf{w}_1, \mathbf{w}_1)}\,|\mathbf{w}_1| = \frac{(\mathbf{w}_1, \mathbf{y})}{|\mathbf{w}_1|}$$
$$= \frac{w_{11}y_1 + \cdots + w_{1n}y_n}{|\mathbf{w}_1|}.$$

Similarly the coordinate for the uth axis is

$$a_u(\mathbf{y})\,|\mathbf{w}_u| = \frac{w_{u1}y_1 + \cdots + w_{un}y_n}{|\mathbf{w}_u|}.$$

(See Figure 10.)

The sum of squares of coordinates of a vector is invariant under an orthogonal transformation. The sum of squares $\Sigma\, y_i^2$ can be expressed as a sum of squares with respect to the new axes:

$$\sum_1^n y_i^2 = \sum_1^r a_u^2(\mathbf{y})\,|\mathbf{w}_u|^2 + s^2(\mathbf{y}).$$

This can be recorded component by component in an *analysis-of-variance* table.

Source	Dimension	Component	Structure of Component
Mean ($\mathbf{w}_1$)	1	$(a_1(\mathbf{y})\,|\mathbf{w}_1|)^2$	$(\alpha_1\,|\mathbf{w}_1| + \sigma a_1(\mathbf{e})\,|\mathbf{w}_1|)^2$
Treatment ($\mathbf{w}_2$)	1	$(a_2(\mathbf{y})\,|\mathbf{w}_2|)^2$	$(\alpha_2\,|\mathbf{w}_2| + \sigma a_2(\mathbf{e})\,|\mathbf{w}_2|)^2$
Variable ($\mathbf{w}_3$)	1	$(a_3(\mathbf{y})\,|\mathbf{w}_3|)^2$	$(\alpha_3\,|\mathbf{w}_3| + \sigma a_3(\mathbf{e})\,|\mathbf{w}_3|)^2$
Residual (**d**)	$n - 3$	$s^2(\mathbf{y})$	$(\sigma s(\mathbf{e}))^2$
Total	n	$\Sigma\, y_i^2$	

The sources for the components are labeled in the manner appropriate to the examples in Section 1.

5 COMPUTATION

The position $[Y]$ is a matrix containing regression coefficients $b_1(\mathbf{y}), \ldots, b_r(\mathbf{y})$ and a residual length $s(\mathbf{y})$. The matrix P, which produces the new orthogonal basis $W = PV$, contains regression coefficients of w-vectors on v-vectors. And the matrix P^{-1}, which produces the new quantities

$$(\alpha_1, \ldots, \alpha_r) = (\beta_1, \ldots, \beta_r)P^{-1}$$

and the new regression coefficients

$$(a_1, \ldots, a_r) = (b_1, \ldots, b_r)P^{-1},$$

also contains regression coefficients, the regression coefficients in fact of v-vectors on w-vectors. All of these regression coefficients can be calculated from the elements of the inner-product matrix YY'. They can all be calculated by a simple repetitive operation applied to that matrix.

The three examples in Section 1 concern a single response vector **y** and its relation to structural vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$: to the first vector; to the first two vectors; to the first three vectors. Correspondingly, in general, there may be interest in a response vector **y** and its relation to structural vectors

$$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r:$$

to the first vector; to the first two vectors; to the first three vectors; . . . . All the regression coefficients for each step can be calculated from the inner products that appear in the matrix YY'. In fact they are obtained as part of applying the repetitive operation to the matrix YY'.

The notation can be extended temporarily to handle the succession of steps by introducing a superscript to indicate the number of structural vectors involved: With r structural vectors the response matrix is $Y^{(r)}$; the

regression coefficients are $b_1^{(r)}(\mathbf{y}), \ldots, b_r^{(r)}(\mathbf{y})$; the residual length is $s^{(r)}(\mathbf{y})$; and so on.

The simple operation can be described most easily by considering two vectors $\mathbf{y}_1, \mathbf{y}_2$ and a single structural vector $\mathbf{v}$. The regression coefficients are

$$b(\mathbf{y}_1) = \frac{(\mathbf{v}, \mathbf{y}_1)}{(\mathbf{v}, \mathbf{v})}, \qquad b(\mathbf{y}_2) = \frac{(\mathbf{v}, \mathbf{y}_2)}{(\mathbf{v}, \mathbf{v})}.$$

A typical inner product of residuals is

$$\begin{aligned}(\mathbf{y}_i - b(\mathbf{y}_i)\mathbf{v}, \mathbf{y}_j - b(\mathbf{y}_j)\mathbf{v}) &= (\mathbf{y}_i, \mathbf{y}_j) - b(\mathbf{y}_i)(\mathbf{v}, \mathbf{y}_j) - b(\mathbf{y}_j)(\mathbf{v}, \mathbf{y}_i) + b(\mathbf{y}_i)b(\mathbf{y}_j)(\mathbf{v}, \mathbf{v}) \\ &= (\mathbf{y}_i, \mathbf{y}_j) - \frac{(\mathbf{v}, \mathbf{y}_i)(\mathbf{v}, \mathbf{y}_j)}{(\mathbf{v}, \mathbf{v})},\end{aligned}$$

(a generalization of this appears in Problems 14, 15, 16); the inner-product matrix for residuals is

$$Q(\mathbf{y}_1, \mathbf{y}_2 : \mathbf{v}) = \begin{bmatrix} (\mathbf{y}_1, \mathbf{y}_1) - \dfrac{(\mathbf{v}, \mathbf{y}_1)(\mathbf{v}, \mathbf{y}_1)}{(\mathbf{v}, \mathbf{v})} & (\mathbf{y}_1, \mathbf{y}_2) - \dfrac{(\mathbf{v}, \mathbf{y}_1)(\mathbf{v}, \mathbf{y}_2)}{(\mathbf{v}, \mathbf{v})} \\ (\mathbf{y}_2, \mathbf{y}_1) - \dfrac{(\mathbf{v}, \mathbf{y}_2)(\mathbf{v}, \mathbf{y}_1)}{(\mathbf{v}, \mathbf{v})} & (\mathbf{y}_2, \mathbf{y}_2) - \dfrac{(\mathbf{v}, \mathbf{y}_2)(\mathbf{v}, \mathbf{y}_2)}{(\mathbf{v}, \mathbf{v})} \end{bmatrix}.$$

Now consider the inner-product matrix for $\mathbf{v}, \mathbf{y}_1, \mathbf{y}_2$:

$$Q(\mathbf{v}, \mathbf{y}_1, \mathbf{y}_2) = \left[\begin{array}{c|cc} (\mathbf{v}, \mathbf{v}) & (\mathbf{v}, \mathbf{y}_1) & (\mathbf{v}, \mathbf{y}_2) \\ \hline (\mathbf{y}_1, \mathbf{v}) & (\mathbf{y}_1, \mathbf{y}_1) & (\mathbf{y}_1, \mathbf{y}_2) \\ (\mathbf{y}_2, \mathbf{v}) & (\mathbf{y}_2, \mathbf{y}_1) & (\mathbf{y}_2, \mathbf{y}_2) \end{array}\right].$$

The simple operation is: *Divide through the first row by the leading element to obtain a new first row; subtract multiples of the new first row from remaining rows to produce zeros in the positions corresponding to the leading element:*

$$\left[\begin{array}{c|cc} 1 & b(\mathbf{y}_1) & b(\mathbf{y}_2) \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$Q(\mathbf{y}_1, \mathbf{y}_2 : \mathbf{v})$}} \\ 0 & & \end{array}\right].$$

The resulting matrix clearly contains the regression coefficients and the inner-product matrix for residuals. The simple operation is equivalent to left multiplication by the triangular matrix

$$\begin{bmatrix} \dfrac{1}{(\mathbf{v}, \mathbf{v})} & 0 & 0 \\ -\dfrac{(\mathbf{y}_1, \mathbf{v})}{(\mathbf{v}, \mathbf{v})} & 1 & 0 \\ -\dfrac{(\mathbf{y}_2, \mathbf{v})}{(\mathbf{v}, \mathbf{v})} & 0 & 1 \end{bmatrix}.$$

Left multiplication by a matrix produces new rows that are linear combinations of old rows.

Now consider the inner product matrix YY':

$$YY' = \begin{bmatrix} (\mathbf{v}_1, \mathbf{v}_1) & \cdots & (\mathbf{v}_1, \mathbf{v}_r) & (\mathbf{v}_1, \mathbf{y}) \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ (\mathbf{v}_r, \mathbf{v}_1) & \cdots & (\mathbf{v}_r, \mathbf{v}_r) & (\mathbf{v}_r, \mathbf{y}) \\ (\mathbf{y}, \mathbf{v}_1) & \cdots & (\mathbf{y}, \mathbf{v}_r) & (\mathbf{y}, \mathbf{y}) \end{bmatrix}.$$

The first r rows contain the matrix array of coefficients for the orthogonality equations for $b_1^{(r)}(\mathbf{y}), \ldots, b_r^{(r)}(\mathbf{y})$. If an equation is multiplied through by a constant, if equations are subtracted, if this operation is repeated so that the coefficients on the left side of the equations become the $r \times r$ identity matrix, then the new "equations" state that the coefficients on the right side are the solutions:

$$\begin{bmatrix} 1 & & 0 & b_1^{(r)}(\mathbf{y}) \\ & \cdot & & \cdot \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & & 1 & b_r^{(r)}(\mathbf{y}) \\ * & \cdots & * & * \end{bmatrix}.$$

In the same manner the first s rows and the first s columns plus the last column give the corresponding matrix array of coefficients for the orthogonality equations for $b_1^{(s)}(\mathbf{y}), \ldots, b_s^{(s)}(\mathbf{y})$. If the preceding operations are applied to the first s rows to produce the $s \times s$ identity, then the "coefficients"

on the right side are the solutions:

$$\begin{bmatrix} 1 & & 0 & * & \cdots & * & b_1^{(s)}(y) \\ & \ddots & \vdots & & & \vdots & \vdots \\ 0 & & 1 & * & \cdots & * & b_s^{(s)}(y) \\ * & \cdots & * & * & \cdots & * & * \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ * & \cdots & * & * & \cdots & * & * \end{bmatrix}.$$

The simple operation can be applied successively. It generates successively the 1×1 identity and the solution $b_1^{(1)}(y)$, the 2×2 identity and the solutions $b_1^{(2)}(y)$, $b_2^{(2)}(y)$, ..., the $r \times r$ identity and the solutions $b_1^{(r)}(y), \ldots, b_r^{(r)}(y)$; and it generates related elements of interest.

For consider the inner product matrix YY' together with an $(r+1) \times (r+1)$ identity matrix:

$$\begin{bmatrix} (v_1, v_1) & \cdots & (v_1, v_r) & (v_1, y) & 1 & & 0 & 0 \\ \vdots & & \vdots & \vdots & & \ddots & & \vdots \\ (v_r, v_1) & \cdots & (v_r, v_r) & (v_r, y) & 0 & & 1 & 0 \\ (y, v_1) & \cdots & (y, v_r) & (y, y) & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Apply the simple operation to the rows of the augmented matrix with the first-row-first-column element as leading element:

$$\left[\begin{array}{c|ccc c|c|ccc} 1 & \boxed{p^{21} & \cdots & p^{r1}} & b_1^{(1)}(y) & (v_1, v_1)^{-1} & 0 & \cdots & 0 \\ \hline 0 & & & & & * & 1 & & 0 \\ \vdots & & & & & \vdots & & \ddots & \\ \vdots & & Q(v_2, \ldots, v_r, y : v_1) & & & \vdots & & & \\ 0 & & & & & * & 0 & & 1 \end{array}\right].$$

The array contains the "solution" $b_1^{(1)}(y)$, the regression coefficient of y on v_1; it contains the matrix Q of inner products of residuals; it contains the inverse of (v_1, v_1), and it contains the regression coefficient $p_{21}(= p^{21})$ of v_2 on v_1. The array also contains the elements $p^{21}, \ldots, p^{r1}$ for the first column of the inverse matrix P^{-1}: the matrix P^{-1} gives the original basis from the orthogonal basis:

$$V = P^{-1}W$$

$$\begin{bmatrix} v_1' \\ \vdots \\ v_r' \end{bmatrix} = \begin{bmatrix} 1 & & & & 0 \\ \boxed{p^{21}} & 1 & & & \\ \vdots & & \ddots & & \\ \boxed{p^{r1}} & & \cdots & p^{r\,r-1} & 1 \end{bmatrix} \begin{bmatrix} w_1' \\ \vdots \\ w_r' \end{bmatrix};$$

the vectors $w_1, \ldots, w_r$ are *orthogonal*; the elements of P^{-1} are *individual* regression coefficients of v vectors on w vectors; the elements $p^{21}, \ldots, p^{r1}$ are the individual regression coefficients of the vectors $v_2, \ldots, v_r$ on the vector $w_1(= v_1)$.

Now apply the simple operation to the rows of the modified array with the second-row-second-column element as leading element:

$$\left[\begin{array}{cc|cccc c|c|ccc} 1 & 0 & \boxed{p_{31}} & * & \cdots & * & b_1^{(2)}(y) & (V^{(2)}V^{(2)\prime})^{-1} & 0 & \cdots & 0 \\ 0 & 1 & \boxed{p^{32}} & p^{42} & \cdots & p^{r2} & b_2^{(2)}(y) & & 0 & \cdots & 0 \\ \hline 0 & 0 & & & & & & * \quad * & 1 & & 0 \\ \vdots & \vdots & & & & & & \vdots \quad \vdots & & \ddots & \\ \vdots & \vdots & & & Q(v_3, \ldots, v_r, y : v_1, v_2) & & & \vdots \quad \vdots & & & \\ 0 & 0 & & & & & & * \quad * & 0 & & 1 \end{array}\right].$$

The array contains the "solution" $b_1^{(2)}(y)$, $b_2^{(2)}(y)$, the regression coefficients of y on v_1, v_2; it contains the matrix Q of inner products of residuals; it contains the regression coefficients p_{31}, $p_{32}(= p^{32})$ of v_3 on v_1, v_2; and it contains the individual regression coefficients $p^{32}, \ldots, p^{r2}$ of $v_3, \ldots, v_r$ on w_2. The array also contains the inverse of the inner product matrix $V^{(2)}V^{(2)\prime}$: the row operations reduce $V^{(2)}V^{(2)\prime}$ to the 2×2 identity; the same row operations applied to the 2×2 identity must produce the inverse matrix.

Now apply the simple operation to the rows of the further modified array with the third-row-third-column element as leading element;

$$\left[\begin{array}{ccc:ccc:c:ccc}
1 & 0 & 0 & \boxed{p_{41}} & * & b_1^{(3)}(\mathbf{y}) & & 0 & \cdots & 0 \\
0 & 1 & 0 & \boxed{p_{42}} & * & b_2^{(3)}(\mathbf{y}) & (V^{(3)}V^{(3)\prime})^{-1} & 0 & \cdots & 0 \\
0 & 0 & 1 & \boxed{p^{43}} & \cdots \; p^{r3} & b_3^{(3)}(\mathbf{y}) & & 0 & \cdots & 0 \\
\hdashline
0 & 0 & 0 & & & & * \;\; * \;\; * & 1 & & 0 \\
\cdot & \cdot & \cdot & & & & \cdot \;\; \cdot \;\; \cdot & & \cdot & \\
\cdot & \cdot & \cdot & & Q(\mathbf{v}_4, \ldots, \mathbf{v}_r, \mathbf{y} : \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) & & \cdot \;\; \cdot \;\; \cdot & & \cdot & \\
\cdot & \cdot & \cdot & & & & \cdot \;\; \cdot \;\; \cdot & & & \cdot \\
0 & 0 & 0 & & & & * \;\; * \;\; * & 0 & & 1
\end{array}\right].$$

The array contains the "solution" $b_1^{(3)}(\mathbf{y})$, $b_2^{(3)}(\mathbf{y})$, $b_3^{(3)}(\mathbf{y})$, the regression coefficients of $\mathbf{y}$ on $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$; it contains the matrix Q of inner products of residuals; it contains the inverse matrix $(V^{(3)}V^{(3)\prime})^{-1}$; it contains the regression coefficients $p_{41}, p_{42}, p_{43}(= p^{43})$ of $\mathbf{v}_4$ on $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$; and it contains the regression coefficients (individual) of $\mathbf{v}_4, \ldots, \mathbf{v}_r$ on $\mathbf{w}_3$.

The simple operation applied r times produces the regression coefficients $b_1^{(r)}(\mathbf{y}), \ldots, b_r^{(r)}(\mathbf{y})$, the inner product matrix of residuals, the inverse matrix $(V^{(r)}V^{(r)\prime})^{-1}$, the elements to give the $(r+1)$st row of P, and the elements to give the rth column of P^{-1}.

Justifications. For $b_1^{(s)}(\mathbf{y}), \ldots, b_s^{(s)}(\mathbf{y})$: the first s columns and the $\mathbf{y}$ column contain the coefficients in the orthogonality equations for the regression coefficients of $\mathbf{y}$ on $\mathbf{v}_1, \ldots, \mathbf{v}_s$; the simple operation applied s times solves these equations; the resulting "equations" present the solutions in the $\mathbf{y}$ column. For $p_{s+11}, \ldots, p_{s+1s}$: the preceding argument with $\mathbf{y}$ replaced by $\mathbf{v}_{s+1}$. For $(V^{(s)}V^{(s)\prime})^{-1}$: the first s rows and s columns contain $V^{(s)}V^{(s)\prime}$, and the first s rows and first s columns after the $\mathbf{y}$ column contain the $s \times s$ identity matrix; the simple operation applied s times amounts to premultiplication by a matrix—the matrix that carries $V^{(s)}V^{(s)\prime}$ into the identity, hence carries the identity into $(V^{(s)}V^{(s)\prime})^{-1}$. For $p^{s+1s}, \ldots, p^{rs}$, and $Q(\mathbf{v}_{s+1}, \ldots, \mathbf{v}_r, \mathbf{y} \colon \mathbf{v}_1, \ldots, \mathbf{v}_s)$ (by induction from s to $s+1$): the matrix $Q(\mathbf{v}_{s+1}, \ldots, \mathbf{v}_r, \mathbf{y} \colon \mathbf{v}_1, \ldots, \mathbf{v}_s)$ is the inner-product matrix for vectors $\mathbf{v} - \Sigma_1^s b_u^{(s)}(\mathbf{v})\mathbf{v}_u$ with $\mathbf{v} = \mathbf{v}_{s+1}, \ldots, \mathbf{v}_r, \mathbf{y}$: the simple operation gives the regression coefficients $c(\mathbf{v})$ for the vectors $\mathbf{v} - \Sigma_1^s b_u^{(s)}(\mathbf{v})\mathbf{v}_u$ (with $\mathbf{v} = \mathbf{v}_{s+2}, \ldots, \mathbf{v}_r, \mathbf{y}$) on the single vector $\mathbf{w}_{s+1} = \mathbf{v}_{s+1} - \Sigma_1^s b_u^{(s)}(\mathbf{v}_{s+1})\mathbf{v}_u$; since $\mathbf{v}_1, \ldots, \mathbf{v}_s$ are orthogonal to $\mathbf{w}_{s+1}$, it follows that the $c(\mathbf{v})$ (with $\mathbf{v} = \mathbf{v}_{s+2}, \ldots, \mathbf{v}_r$) are also the regression coefficients $p^{s+1s}, \ldots, p^{rs}$ *of the vectors* $\mathbf{v}$ (with $\mathbf{v} = \mathbf{v}_{s+2}, \ldots, \mathbf{v}_r$) on the single vector $\mathbf{w}_{s+1}$. The simple operation gives the inner-product matrix of the residuals $\mathbf{v} - \Sigma_1^s b_u^{(s)}(\mathbf{v})\mathbf{v}_u - c(\mathbf{v})(\mathbf{v}_{s+1} - \Sigma_1^s b_u^{(s)}(\mathbf{v}_{s+1})\mathbf{v}_u)$; such a residual has the form $\mathbf{v}$ minus a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_s, \mathbf{v}_{s+1}$, and it is orthogonal to $\mathbf{v}_1, \ldots, \mathbf{v}_s, \mathbf{w}_{s+1}$, hence to $\mathbf{v}_1, \ldots, \mathbf{v}_s, \mathbf{v}_{s+1}$; such a residual must then be the residual of $\mathbf{v}$ orthogonalized to $\mathbf{v}_1, \ldots, \mathbf{v}_s, \mathbf{v}_{s+1}$, and the inner-product matrix must then be the inner-product matrix of such residuals.

6 THE EXAMPLES

Consider the examples in Section 1, and suppose that the response vector $\mathbf{y}$ is given by the final row in the augmented matrix

$$Y = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\
7 & 12 & 8 & 15 & 13 & 14 & 17 & 14 & 14 & 17 & 14 & 17
\end{bmatrix}.$$

In this example the structural vectors have a simple form that allows the orthogonal vectors to be written down by inspection and the regression coefficients to be calculated as averages. This simple form also allows the examples to serve as a transparent first illustration of the computation methods in Section 5.

The inner product matrix with appended identity matrix is

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{y}$				
$\mathbf{v}_1$	$\underline{12}$	9	9	162	1	0	0	0
$\mathbf{v}_2$	9	9	9	135	0	1	0	0
$\mathbf{v}_3$	9	9	15	141	0	0	1	0
$\mathbf{y}$	162	135	141	2302	0	0	0	1.

A first application of the simple operation produces the numbers appropriate to the regression model† with one structural vector:

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{y}$				
$\mathbf{v}_1$	1	$\boxed{\frac{3}{4}}$	$\frac{3}{4}$	$\boxed{13\frac{1}{2}}$	$\boxed{\frac{1}{12}}$	0	0	0
$\mathbf{v}_2$	0	$\underline{2\frac{1}{4}}$	$2\frac{1}{4}$	$13\frac{1}{2}$	$-\frac{3}{4}$	1	0	0
$\mathbf{v}_3$	0	$2\frac{1}{4}$	$8\frac{1}{4}$	$19\frac{1}{2}$	$-\frac{3}{4}$	0	1	0
$\mathbf{y}$	0	$13\frac{1}{2}$	$19\frac{1}{2}$	$\boxed{115}$	$-13\frac{1}{2}$	0	0	1.

† The model in this case is also a measurement model but with residual length $s^{(1)}(\mathbf{y})$ replacing standard deviation s_y.

The position and the reference point are given in the expressions

$$Y^{(1)} = [Y^{(1)}]\,D(Y^{(1)}) = \begin{pmatrix} 1 & 0 \\ 13\tfrac{1}{2} & \sqrt{115} \end{pmatrix} D(Y^{(1)}),$$

$$D(Y^{(1)}) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{-6\frac{1}{2}}{\sqrt{115}} & \frac{-1\frac{1}{2}}{\sqrt{115}} & \frac{-5\frac{1}{2}}{\sqrt{115}} & \frac{1\frac{1}{2}}{\sqrt{115}} & \frac{-\frac{1}{2}}{\sqrt{115}} & \frac{\frac{1}{2}}{\sqrt{115}} & \frac{3\frac{1}{2}}{\sqrt{115}} & \frac{\frac{1}{2}}{\sqrt{115}} & \frac{\frac{1}{2}}{\sqrt{115}} & \frac{3\frac{1}{2}}{\sqrt{115}} & \frac{\frac{1}{2}}{\sqrt{115}} & \frac{3\frac{1}{2}}{\sqrt{115}} \end{pmatrix}$$

The analysis-of-variance table can be calculated:

Source	Dimension	Component	Structure of Component
Mean ($\mathbf{w}_1$)	1	2187	$(\alpha_1\sqrt{12} + \sigma a_1(\mathbf{e})\sqrt{12})^2$
Residual ($\mathbf{d}^{(1)}$)	11	115	$(\sigma s^{(1)}(\mathbf{e}))^2$
	12	2302	

The component 2187 is the difference between the original squared length $|\mathbf{y}|^2 = \sum y_i^2 = 2302$ and the squared residual length $(s^{(1)}(\mathbf{y}))^2 = 115$. Some elements in the full matrices, P, P^{-1}, are available:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \boxed{-\frac{3}{4}} & 1 & 0 \\ & & 1 \end{pmatrix}, \qquad P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{4} & 1 & 0 \\ \frac{3}{4} & & 1 \end{pmatrix}.$$

A second application of the simple operation produces the numbers appropriate to the regression model with two structural vectors:

$$\begin{array}{c|cccc|cccc} & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{y} & & & & \\ \hline \mathbf{v}_1 & 1 & 0 & 0 & 9 & \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ \mathbf{v}_2 & 0 & 1 & \boxed{1} & 6 & -\frac{1}{3} & \frac{4}{9} & 0 & 0 \\ \mathbf{v}_3 & 0 & 0 & \underline{6} & 6 & 0 & -1 & 1 & 0 \\ \mathbf{y} & 0 & 0 & 6 & \boxed{34} & -9 & -6 & 0 & 1. \end{array}$$

The position and the reference point are given in the expressions

$$Y^{(2)} = [Y^{(2)}]D(Y^{(2)}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 9 & 6 & \sqrt{34} \end{pmatrix} D(Y^{(2)}),$$

$$D(Y^{(2)}) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{-2}{\sqrt{34}} & \frac{3}{\sqrt{34}} & \frac{-1}{\sqrt{34}} & 0 & \frac{-2}{\sqrt{34}} & \frac{-1}{\sqrt{34}} & \frac{2}{\sqrt{34}} & \frac{-1}{\sqrt{34}} & \frac{-1}{\sqrt{34}} & \frac{2}{\sqrt{34}} & \frac{-1}{\sqrt{34}} & \frac{2}{\sqrt{34}} \end{pmatrix}$$

$$W^{(2)} = P^{(2)}V^{(2)} = \begin{pmatrix} 1 & 0 \\ -\frac{3}{4} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

The analysis-of-variance table can be calculated:

Source	Dimension	Component	Structure of Component
Mean ($\mathbf{w}_1$)	1	2187	$(\alpha_1\sqrt{12} + \sigma a_1(\mathbf{e})\sqrt{12})^2$
Treatment ($\mathbf{w}_2$)	1	81	$(\alpha_2\sqrt{2\frac{1}{4}} + \sigma a_2(\mathbf{e})\sqrt{2\frac{1}{4}})^2$
Residual ($\mathbf{d}^{(2)}$)	10	34	$(\sigma s^{(2)}(\mathbf{e}))^2$
	12	2302	

The component 81 is the difference between the squared residual length $(s^{(1)}(\mathbf{y}))^2 = 115$ relative to $\mathbf{v}_1$ and the squared residual length $(s^{(2)}(\mathbf{y}))^2 = 34$ relative to $\mathbf{v}_1$ and $\mathbf{v}_2$. The remaining elements in the matrices P and P^{-1} are available:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 \\ \boxed{0} & \boxed{-1} & 1 \end{pmatrix}, \qquad P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{4} & 1 & 0 \\ \frac{3}{4} & \boxed{1} & 1 \end{pmatrix}.$$

A third application of the simple operation produces the numbers appropriate to the regression model with three structural vectors:

$$\begin{array}{c|cccc|cccc} & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{y} & & & & \\ \hline \mathbf{v}_1 & 1 & 0 & 0 & 9 & \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ \mathbf{v}_2 & 0 & 1 & 0 & 5 & -\frac{1}{3} & \frac{11}{18} & -\frac{1}{6} & 0 \\ \mathbf{v}_3 & 0 & 0 & 1 & 1 & 0 & -\frac{1}{6} & \frac{1}{6} & 0 \\ \mathbf{y} & 0 & 0 & 0 & \boxed{28} & -9 & -5 & -1 & 1. \end{array}$$

The position and the reference point are given in the expressions

$$Y = [Y]D(Y) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 9 & 5 & 1 & \sqrt{28} \end{bmatrix} D(Y),$$

$$D(Y) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ \frac{-2}{\sqrt{28}} & \frac{3}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{1}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{0}{\sqrt{28}} & \frac{2}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{1}{\sqrt{28}} & \frac{-2}{\sqrt{28}} & \frac{1}{\sqrt{28}} \end{bmatrix},$$

$$W = PV = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \mathbf{v}_3' \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

The analysis-of-variance table can be calculated:

Source	Dimension	Component	Structure of Component
Mean ($\mathbf{w}_1$)	1	2187	$(\alpha_1\sqrt{12} + \sigma a_1(\mathbf{e})\sqrt{12})^2$
Treatment ($\mathbf{w}_2$)	1	81	$(\alpha_2\sqrt{2\frac{1}{4}} + \sigma a_2(\mathbf{e})\sqrt{2\frac{1}{4}})^2$
Variable ($\mathbf{w}_3$)	1	6	$(\alpha_3\sqrt{6} + \sigma a_3(\mathbf{e})\sqrt{6})^2$
Residual ($\mathbf{d}$)	9	28	$(\sigma^2 s(\mathbf{e}))^2$
	12	2302	

The component 6 is the difference between the squared residual length $(s^{(2)}(\mathbf{y}))^2 = 34$ and the squared residual length $(s(\mathbf{y}))^2 = 28$.

The numbers are also available for the analysis in terms of the orthogonal basis. With three structural vectors, the position and reference point are

given in the expressions

$$\bar{Y} = [\bar{Y}]D(\bar{Y}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 13\frac{1}{2} & 6 & 1 & \sqrt{28} \end{bmatrix} D(\bar{Y}),$$

$$D(\bar{Y}) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{4} & -\frac{3}{4} & -\frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ \frac{-2}{\sqrt{28}} & \frac{3}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{1}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{0}{\sqrt{28}} & \frac{2}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{1}{\sqrt{28}} & \frac{-2}{\sqrt{28}} & \frac{1}{\sqrt{28}} \end{bmatrix}.$$

The regression coefficients for $[\bar{Y}]$ are circled in the three matrix arrays.

The computations on the matrix Y can be used to illustrate a matrix factorization that will appear in later sections of this chapter. The matrix Y has been factored:

$$Y = [Y]D(Y) = [Y]\begin{bmatrix} P & 0 \\ 0 & 1 \end{bmatrix}^{-1} D(\bar{Y})$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 9 & 5 & 1 & \sqrt{28} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{3}{4} & 1 & 0 & 0 \\ \frac{3}{4} & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{w}_1' \\ \mathbf{w}_2' \\ \mathbf{w}_3' \\ \mathbf{d}' \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{3}{4} & 1 & 0 & 0 \\ \frac{3}{4} & 1 & 1 & 0 \\ 13\frac{1}{2} & 6 & 1 & \sqrt{28} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1' \\ \mathbf{w}_2' \\ \mathbf{w}_3' \\ \mathbf{d}' \end{bmatrix}.$$

The vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{d}$ are mutually *orthogonal.* Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{d}$ be the corresponding unit or *normal* vectors:

$$\mathbf{u}_1 = \frac{\mathbf{w}_1}{|\mathbf{w}_1|}, \quad \mathbf{u}_2 = \frac{\mathbf{w}_2}{|\mathbf{w}_2|}, \quad \mathbf{u}_3 = \frac{\mathbf{w}_3}{|\mathbf{w}_3|};$$

the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{d}$ form an *orthonormal* set (the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ are unit vectors for the new axes mentioned at the end of Section 4).

$$Y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{3}{4} & 1 & 0 & 0 \\ \frac{3}{4} & 1 & 1 & 0 \\ 13\frac{1}{2} & 6 & 1 & \sqrt{28} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 & 0 \\ 0 & \sqrt{2\frac{1}{4}} & 0 & 0 \\ 0 & 0 & \sqrt{6} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1' \\ \mathbf{u}_2' \\ \mathbf{u}_3' \\ \mathbf{d}' \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{12} & 0 & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & \sqrt{6} & 0 \\ 13\frac{1}{2}\sqrt{12} & 6\sqrt{2\frac{1}{4}} & \sqrt{6} & \sqrt{28} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1' \\ \mathbf{u}_2' \\ \mathbf{u}_3' \\ \mathbf{d}' \end{bmatrix};$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 7 & 12 & 8 & 15 & 13 & 14 & 17 & 14 & 14 & 17 & 14 & 17 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{12} & 0 & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & \sqrt{6} & 0 \\ 13\frac{1}{2}\sqrt{12} & 6\sqrt{2\frac{1}{4}} & \sqrt{6} & \sqrt{28} \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ \frac{-\frac{3}{4}}{\sqrt{2\frac{1}{4}}} & \frac{-\frac{3}{4}}{\sqrt{2\frac{1}{4}}} & \frac{-\frac{3}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} & \frac{\frac{1}{4}}{\sqrt{2\frac{1}{4}}} \\ 0 & 0 & 0 & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & 0 & 0 & 0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{28}} & \frac{3}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{1}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{0}{\sqrt{28}} & \frac{2}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{-1}{\sqrt{28}} & \frac{1}{\sqrt{28}} & \frac{-2}{\sqrt{28}} & \frac{1}{\sqrt{28}} \end{bmatrix}$$

The original matrix Y has been factored into a *positive lower triangular* matrix and *a matrix with four orthonormal vectors. The four row vectors in the original matrix can be represented by the four row vectors in the triangular matrix: These row vectors in the triangular matrix are with respect to axes given by the orthonormal vectors in the final matrix.* The inner product matrices in the two representations are of course equal:

$$YY' = \begin{bmatrix} \sqrt{12} & 0 & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & \sqrt{6} & 0 \\ 13\frac{1}{2}\sqrt{12} & 6\sqrt{2\frac{1}{4}} & \sqrt{6} & \sqrt{28} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & 0 & 0 \\ \frac{3}{4}\sqrt{12} & \sqrt{2\frac{1}{4}} & \sqrt{6} & 0 \\ 13\frac{1}{2}\sqrt{12} & 6\sqrt{2\frac{1}{4}} & \sqrt{6} & \sqrt{28} \end{bmatrix}',$$

a triangular square-root factorization of YY'.

7 THE REDUCED MODEL

Consider the regression model as developed in Sections 2, 3:

$$f(E)\, dE,$$

$$[Y] = \theta[E], \qquad D(Y) = D(E).$$

Let $[Y]$ be the transformation variable defined in Section 3 and $W = PV$ be the orthogonal basis defined in Section 4.

The invariant differential on the space $\mathfrak{X} = R^n$ can be obtained (Section 3, Chapter Two) from the Jacobian of a transformation on that space:

$$\tilde{y}_1 = a_1 v_{11} + \cdots + a_r v_{r1} + c y_1$$

$$\vdots$$

$$\tilde{y}_n = a_1 v_{1n} + \cdots + a_r v_{rn} + c y_n,$$

$$J_n(g; \mathbf{y}) = c^n,$$

$$J_n(Y) = s^n(\mathbf{y}),$$

$$dm(Y) = \frac{\prod dy_i}{s^n(\mathbf{y})} = \frac{dY}{s^n(\mathbf{y})}$$

The left and right invariant differentials on the group can be obtained (Sections 3 and 4 in Chapter Two) from Jacobians of transformations on the

group:

$$\begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ \tilde{a}_1 & \cdots & & \tilde{a}_r & \tilde{c} \end{bmatrix} = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ a_1 & \cdots & & a_r & c \end{bmatrix} \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ a_1^* & \cdots & & a_r^* & c^* \end{bmatrix},$$

$$\begin{aligned} \tilde{a}_1 &= a_1 + ca_1^* \\ &\cdot \\ &\cdot \\ &\cdot \\ \tilde{a}_r &= a_r + ca_r^*, \\ \tilde{c} &= \qquad cc^*. \end{aligned}$$

$$J_{r+1} = \begin{vmatrix} c & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & c & 0 \\ 0 & \cdots & & 0 & c \end{vmatrix} = c^{r+1}, \qquad J_{r+1}^* = \begin{vmatrix} 1 & & & 0 & a_1^* \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & a_r^* \\ 0 & \cdots & & 0 & c^* \end{vmatrix} = c^*.$$

$$J_{r+1}(g) = c^{r+1}, \qquad J_{r+1}^*(g) = c,$$

$$d\mu(g) = \frac{\prod da_u\, dc}{c^{r+1}} = \frac{dg}{c^{r+1}}, \qquad d\nu(g) = \frac{\prod da_u\, dc}{c} = \frac{dg}{c},$$

$$\Delta(g) = \frac{c}{c^{r+1}} = \frac{1}{c^r}.$$

7.1 General Distributions. The conditional probability element for the error position $[E]$ given the orbit $D(E) = D$ can be obtained by exchanging invariant differentials (Section 4 in Chapter Two):

$$\begin{aligned} g([E]:D)\, d[E] &= k(D) f([E]D) \frac{s^n}{s^{r+1}}\, d[E] \\ &= k(D) \prod_{i=1}^{n} f\Big(\sum_u b_u v_{ui} + s d_i\Big) s^{n-r-1} \prod db_u\, ds. \end{aligned}$$

And the structural probability element for θ given Y can be obtained by manipulating invariant differentials (Section 5 in Chapter Two):

$$\begin{aligned} g^*(\theta:Y)\, d\theta &= k(D(Y)) f(\theta^{-1}Y) \left(\frac{s(y)}{\sigma}\right)^n (s(y))^{-r}\, d\nu(\theta) \\ &= k(D(Y)) \prod_1^n f\left(\frac{y_i - \sum \beta_u v_{ui}}{\sigma}\right) \left(\frac{s(y)}{\sigma}\right)^n (s(y))^{-r} \frac{\prod d\beta_u\, d\sigma}{\sigma}. \end{aligned}$$

7.2 Decomposition. The structural equation for the regression model,

$$\begin{aligned} b_1(y) &= \beta_1 + \sigma b_1(e) \\ &\cdot \\ &\cdot \\ &\cdot \\ b_r(y) &= \beta_r + \sigma b_r(e), \\ s(y) &= \qquad \sigma s(e), \end{aligned}$$

can be separated into a part concerning the β's:

$$\begin{aligned} \frac{b_1(y) - \beta_1}{s(y)} &= \frac{b_1(e)}{s(e)} = t_1(e) \\ &\cdot \\ &\cdot \\ &\cdot \\ \frac{b_r(y) - \beta_r}{s(y)} &= \frac{b_r(e)}{s(e)} = t_r(e); \end{aligned}$$

and a part concerning σ:

$$\frac{s(y)}{\sigma} = s(e).$$

In a related manner the regression-scale group has a *location subgroup*:

$$L = \left\{ \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ a_1 & \cdots & & a_r & 1 \end{bmatrix} : \quad -\infty < a_u < \infty \right\};$$

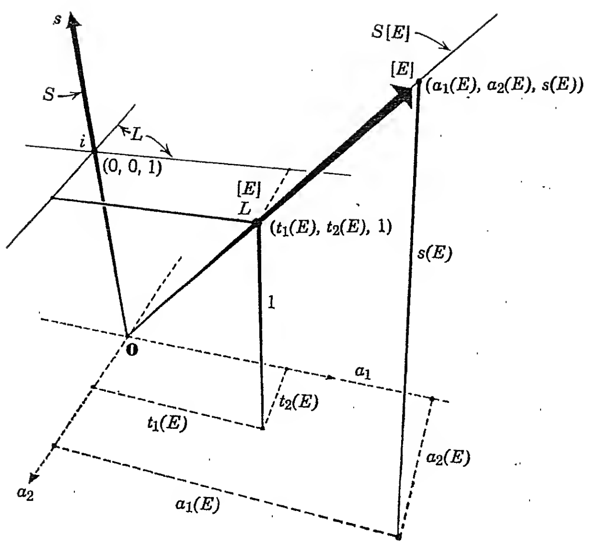


Figure 11 The regression-scale group as G^*, recording the position $[E]$ of E on its orbit. The *normal* regression subgroup L. The scale subgroup S; and an orbit or right coset $S[E]$ of the scale subgroup S.

and a *scale subgroup*:

$$S = \left\{ \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ 0 & \cdots & & 0 & c \end{bmatrix} : \ 0 < c < \infty \right\}.$$

(See Figure 11.)

7.3 **Location Distributions.** Consider inference concerning the β's. The marginal distribution of the error variable $(t_1(\mathrm{e}), \ldots, t_r(\mathrm{e}))$ can be obtained from the error probability distribution

$$k(D)f([E]D)s^{n-r-1} \prod db_u\, ds:$$

the joint distribution of $(t_1, \ldots, t_r, s)$ is

$$k(D)f([E]D)s^{n-1} \prod dt_u\, ds;$$

the marginal distribution of $(t_1(\mathrm{e}), \ldots, t_r(\mathrm{e}))$ is

$$g_L(\mathbf{t}: D) \prod dt_u = k(D) \int_0^\infty \prod_1^n f\left(s\left(\sum_u t_u v_{ui} + d_i\right)\right) s^{n-1}\, ds \cdot \prod dt_u.$$

The structural distribution for $(\beta_1, \ldots, \beta_r)$ is

$$g_L^*(\boldsymbol{\beta}: Y) \prod d\beta_u$$
$$= \frac{k(D)}{s^r(\mathrm{y})} \int_0^\infty \prod_1^n f\left(s\left(\sum_u \frac{b_u(\mathrm{y}) - \beta_u}{s(\mathrm{y})} v_{ui} + d_i\right)\right) s^{n-1}\, ds \cdot \prod d\beta_u.$$

A group element g can be factored uniquely,

$$g = [g]_S[g]_L = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ 0 & \cdots & & 0 & c \end{bmatrix} \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ \frac{a_1}{c} & \cdots & & \frac{a_r}{c} & 1 \end{bmatrix},$$

as an element of S times an element of L (G is a semidirect product of S and L; definition on *p.* 69) The scale group S generates orbits (right cosets) on the error space G^*:

$$S \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ b_1(\mathrm{e}) & \cdots & & b_r(\mathrm{e}) & s(\mathrm{e}) \end{bmatrix} = S \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ t_1(\mathrm{e}) & \cdots & & t_r(\mathrm{e}) & 1 \end{bmatrix};$$

see Figure 11. The error *variable* $(t_1(\mathrm{e}), \ldots, t_r(\mathrm{e}))$ indexes these scale-group orbits. A group element g can alternatively be factored uniquely:

$$g = [g]_L[g]_S = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ a_1 & \cdots & & a_r & 1 \end{bmatrix} \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ 0 & \cdots & & 0 & c \end{bmatrix}.$$

as an element of L times an element of S. The scale group S generates the left coset

$$\theta S = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ \beta_1 & \cdots & & \beta_r & \sigma \end{bmatrix} S = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ \beta_1 & \cdots & & \beta_r & 1 \end{bmatrix} S,$$

for the quantity θ. Thus *the right coset* distribution of the t's produces a *left coset* distribution for the β's.

7.4 Scale Distributions. Now consider inference concerning σ. The marginal distribution of the error quantity $s(\mathbf{e})$ is

$$g_S(s: D)\, ds = k(D) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_1^n f\left(\sum_u b_u v_{ui} + s d_i\right) \prod db_u \cdot s^{n-r-1}\, ds.$$

The structural distribution for σ is

$$g_S^*(\sigma : Y)\, d\sigma = k(D) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_1^n f\left(\sum_u b_u v_{ui} + \frac{s(\mathbf{y})}{\sigma} d_i\right) \prod db_u \cdot \left(\frac{s(\mathbf{y})}{\sigma}\right)^{n-r} \frac{d\sigma}{\sigma}.$$

The location group L generates orbits (right cosets) on the error space G^*:

$$L[E] = L \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ 0 & \cdots & & 0 & s(\mathbf{e}) \end{bmatrix};$$

and correspondingly generates a left coset

$$\theta L = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ 0 & \cdots & & 0 & \sigma \end{bmatrix} L$$

for the quantity θ. *The right coset* distribution of $s(\mathbf{e})$ produces *a left coset* distribution of σ.

7.5 Tests of Location. Consider tests of significance concerning the location quantity $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$. Suppose some outside source has indicated that

$$\beta_1 = \beta_{10}, \ldots, \beta_r = \beta_{r0}.$$

The *hypothesis* $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, together with the structural equation, leads to the value

$$(t_1(\mathbf{e}), \ldots, t_r(\mathbf{e})) = \left(\frac{b_1(\mathbf{y}) - \beta_{10}}{s(\mathbf{y})}, \ldots, \frac{b_r(\mathbf{y}) - \beta_{r0}}{s(\mathbf{y})}\right)$$

for a characteristic of the unknown error $\mathbf{e}$. This value can be compared with the distribution

$$g_L(\mathbf{t} : D(Y)) \prod dt_u$$

derived from the error probability distribution; and the hypothesis can be assessed accordingly.

Now suppose some outside source has indicated that $\beta_r = \beta_{r0}$. The *hypothesis* $\beta_r = \beta_{r0}$, together with the structural equation, gives the value

$$t_r(\mathbf{e}) = \frac{b_r(\mathbf{y}) - \beta_{r0}}{s(\mathbf{y})} = \frac{a_r(\mathbf{y}) - \beta_{r0}}{s(\mathbf{y})}$$

for a characteristic of the unknown error $\mathbf{e}$ (the coefficient of the last structural vector is unaffected by the shift to the orthogonal basis). This value can be compared with the distribution of the variable

$$t_r(\mathbf{e}) = \frac{b_r(\mathbf{e})}{s(\mathbf{e})} = \frac{a_r(\mathbf{e})}{s(\mathbf{e})}$$

derived from the distribution $g([E]: D(Y))\, d[E]$ or equivalently from the distribution $g_L(\mathbf{t}: D(Y))d\mathbf{t}$ and the hypothesis can be assessed accordingly.

7.6 A Test of Scale. Now consider tests of significance concerning the scale quantity σ. Suppose some outside source has indicated that $\sigma = \sigma_0$. The hypothesis $\sigma = \sigma_0$ leads to the value

$$s(\mathbf{e}) = \frac{s(\mathbf{y})}{\sigma_0}$$

for a characteristic of the unknown error $\mathbf{e}$. This value for $s(\mathbf{e})$ can be compared with the distribution

$$g_S(s: D(Y))\, ds,$$

and the hypothesis assessed accordingly.

8 WITH NORMAL ERROR

Consider the regression model with standard normal error variables:

$$f(E)\,dE = (2\pi)^{-n/2} \exp\{-\tfrac{1}{2}|\mathbf{e}|^2\} \prod de_i,$$
$$[Y] = \theta[E], \qquad D(Y) = D(E).$$

The alternative model with orthogonal basis is

$$f(\bar{E})\,d\bar{E} = (2\pi)^{-n/2} \exp\{-\tfrac{1}{2}|\mathbf{e}|^2\} \prod de_i,$$
$$[\bar{Y}] = \bar{\theta}[\bar{E}], \qquad D(\bar{Y}) = D(\bar{E}).$$

8.1 General Distributions: Error. The error probability distribution (*orthogonal* basis) is

$$g([\bar{E}]:\bar{D})\,d[\bar{E}]$$
$$= k(\bar{D})(2\pi)^{-n/2} \exp\{-\tfrac{1}{2}|\textstyle\sum a_u \mathbf{w}_u + s\mathbf{d}|^2\} s^{n-r-1} \prod da_u\,ds$$
$$= k'(\bar{D}) \exp\{-\tfrac{1}{2}(\textstyle\sum a_u^2 |\mathbf{w}_u|^2 + s^2)\} s^{n-r-1} \prod da_u\,ds$$
$$= \frac{\prod_u |\mathbf{w}_u|}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}\textstyle\sum a_u^2 |\mathbf{w}_u|^2\} \prod da_u$$
$$\cdot \frac{1}{\Gamma((n-r)/2)} \left(\frac{s^2}{2}\right)^{(n-r)/2-1} \exp\left\{-\frac{s^2}{2}\right\} d\frac{s^2}{2}$$
$$= \frac{|WW'|^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}\mathbf{a}'WW'\mathbf{a}\}\,d\mathbf{a} \cdot \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} s^{n-r-1} \exp\left\{-\frac{s^2}{2}\right\} ds.$$

The error regression coefficients $a_1(\mathbf{e}), \ldots, a_r(\mathbf{e})$ are independent normal variables with means equal to zero and with variances $|\mathbf{w}_1|^{-2}, \ldots, |\mathbf{w}_r|^{-2}$; the squared residual length $s^2(\mathbf{e})$ has a chi-square distribution on $n - r$ degrees of freedom; $\mathbf{a}(\mathbf{e})$ and $s(\mathbf{e})$ are statistically independent; the error probability distribution does not depend on the orbit as given by $\bar{D}$. It follows then that the elements $a_1^2(\mathbf{e})\,|\mathbf{w}_1|^2, \ldots, a_r^2(\mathbf{e})\,|\mathbf{w}_r|^2$, $s^2(\mathbf{e})$ in the structure-of-component column of the analysis-of-variance table (Section 4) are independent chi-square variables with degrees of freedom $1, \ldots, 1$, $n - r$ as given in the dimension column. Note: The density factored so that the variables separated; the factors have normal and chi-square form; the usual normal and chi-square normalizing constants were introduced.

The error probability distribution (*original* basis) can be obtained by the change of variable:

$$(a_1, \ldots, a_r) = (b_1, \ldots, b_r)P^{-1},$$
$$s = s,$$

and the substitution $W = PV$:

$$g([E]:D)\,d[E]$$
$$= \frac{|VV'|^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}\mathbf{b}'VV'\mathbf{b}\}\,d\mathbf{b} \cdot \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} s^{n-r-1} \exp\left\{-\frac{s^2}{2}\right\} ds.$$

The error regression coefficients $b_1(\mathbf{e}), \ldots, b_r(\mathbf{e})$ have a multivariate normal distribution† *with means 0 and covariance matrix $(VV')^{-1}$; the squared residual length $s^2(\mathbf{e})$ has a chi-square distribution on $n - r$ degrees of freedom; $\mathbf{b}(\mathbf{e})$, $s(\mathbf{e})$ are statistically independent.*

8.2 General Distributions: Structural. The structural probability element (*orthogonal* basis) can be obtained from the error distribution by substitution, or from the expression in the preceding subsection:

$$g^*(\bar{\theta}:\bar{Y})\,d\bar{\theta} = \frac{|WW'|^{1/2}}{(2\pi\sigma^2)^{r/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{\alpha} - \mathbf{a}(\mathbf{y}))' \frac{WW'}{\sigma^2} (\boldsymbol{\alpha} - \mathbf{a}(\mathbf{y}))\right\} d\boldsymbol{\alpha}$$
$$\cdot \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} \left(\frac{s(\mathbf{y})}{\sigma}\right)^{n-r-1} \exp\left\{-\frac{s^2(\mathbf{y})}{2\sigma^2}\right\} \frac{s(\mathbf{y})}{\sigma} \frac{d\sigma}{\sigma}.$$

The quantities $\alpha_1, \ldots, \alpha_r$ conditional on σ are independent normal variables with means $a_1(\mathbf{y}), \ldots, a_r(\mathbf{y})$ and variances $|\mathbf{w}_1|^{-2}\sigma^2, \ldots, |\mathbf{w}_r|^{-2}\sigma^2$; the marginal distribution of σ^2 is that of $s^2(\mathbf{y})\chi^{-2}$, where χ^2 has a chi-square distribution on $n - r$ degrees of freedom.

The structural probability element (*original* basis) is

$$g^*(\theta:Y)\,d\theta = \frac{|VV'|^{1/2}}{(2\pi\sigma^2)^{r/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{\beta} - \mathbf{b}(\mathbf{y}))' \frac{VV'}{\sigma^2} (\boldsymbol{\beta} - \mathbf{b}(\mathbf{y}))\right\} d\boldsymbol{\beta}$$
$$\cdot \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} \left(\frac{s(\mathbf{y})}{\sigma}\right)^{n-r-1} \exp\left\{-\frac{s^2(\mathbf{y})}{2\sigma^2}\right\} \frac{s(\mathbf{y})}{\sigma} \frac{d\sigma}{\sigma}.$$

The quantity $(\beta_1, \ldots, \beta_r)$ conditional on σ is multivariate normal with mean $(b_1(\mathbf{y}), \ldots, b_r(\mathbf{y}))$ and covariance matrix $(VV')^{-1}\sigma^2$; the marginal distribution of σ^2 is that of $s^2(\mathbf{y})\chi^{-2}$, where χ^2 has a chi-square distribution on $n - r$ degrees of freedom.

8.3 Location Distributions: Error. The marginal probability distribution (*orthogonal* basis) of the error variable $\bar{\mathbf{t}} = \bar{\mathbf{t}}(\mathbf{e})$,

$$\bar{t}_1(\mathbf{e}) = \frac{a_1(\mathbf{e})}{s(\mathbf{e})}, \ldots, \bar{t}_r(\mathbf{e}) = \frac{a_r(\mathbf{e})}{s(\mathbf{e})},$$

† A nonsingular linear transformation of a vector of independent normal variables gives a *multivariate normal variable*; e.g., $\mathbf{b}' = \mathbf{a}'P$, $\mathbf{a}' = \mathbf{b}'P^{-1}$. Note that $E(\mathbf{a}) = \mathbf{0}$, $E(a_u^2) = (\mathbf{w}_u, \mathbf{w}_u)^{-1}$, $E(a_u a_v) = 0\ (u \neq v)$, $E(\mathbf{a}\mathbf{a}') = (WW')^{-1}$; hence that $E(P'^{-1}\mathbf{b}) = \mathbf{0}$, $E(P'^{-1}\mathbf{b}\mathbf{b}'P^{-1}) = (WW')^{-1}$; and hence that $E(\mathbf{b}) = \mathbf{0}$, $E(\mathbf{b}\mathbf{b}') = (VV')^{-1}$.

can be obtained from the joint distribution of $\mathbf{a}, s$ by substitution and integration:

$$\int_0^\infty \frac{|WW'|^{1/2}}{(2\pi)^{r/2}} \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} \exp\left\{-\frac{s^2}{2}(1 + \bar{\mathbf{t}}'WW'\bar{\mathbf{t}})\right\} s^{n-1}\, ds\, d\bar{\mathbf{t}}$$

$$= \frac{|WW'|^{1/2} A_{n-r}}{(1 + \bar{\mathbf{t}}'WW'\bar{\mathbf{t}})^{n/2}}\, d\bar{\mathbf{t}} \int_0^\infty \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{u^2}{2}\right\} u^{n-1} du$$

$$= \frac{A_{n-r}}{A_n} \frac{|WW'|^{1/2}}{(1 + \bar{\mathbf{t}}'WW'\bar{\mathbf{t}})^{n/2}}\, d\bar{\mathbf{t}}.$$

This is an r-dimensional analog of the t-distribution: The a's are independent normal variables with means equal to zero; the s is chi on $n - r$; the vector $\bar{\mathbf{t}}$ is the vector of a's expressed in units of s. The distribution in the case $r = 1$ and $|\mathbf{w}_1| = 1$ is

$$\frac{A_{n-1}}{A_n} \frac{1}{(1 + \bar{t}^2)^{n/2}}\, d\bar{t};$$

the variable $\bar{t}$ is a *simplified t-variable;* it is an ordinary t-variable divided by the square root of its degrees of freedom. The marginal distribution of

$$\bar{t}_r(\mathbf{e}) = \frac{b_r(\mathbf{e})}{s(\mathbf{e})} = \frac{a_r(\mathbf{e})}{s(\mathbf{e})}$$

can then be obtained by noting that $\bar{t}$ is a central normal variable in units of a chi variable; the marginal distribution is

$$\frac{A_{n-r}}{A_{n-r+1}} \frac{|\mathbf{w}_r|}{(1 + \bar{t}_r^2 |\mathbf{w}_r|^2)^{(n-r+1)/2}}\, d\bar{t}_r;$$

it is a rescaled t-distribution on $n - r$ degrees of freedom. The distribution of $\bar{\mathbf{t}}$ for general r and with $|\mathbf{w}_1| = \cdots = |\mathbf{w}_r| = 1$ is

$$\frac{A_{n-r}}{A_n} \frac{1}{(1 + \sum \bar{t}_u^2)^{n/2}}\, d\bar{\mathbf{t}};$$

this is the *simplified t-distribution in r-dimensions with $n - r$ degrees of freedom.*

The marginal probability distribution (*original* basis) of the error variable $\mathbf{t}(\mathbf{e})$,

$$t_1(\mathbf{e}) = \frac{b_1(\mathbf{e})}{s(\mathbf{e})}, \ldots, t_r(\mathbf{e}) = \frac{b_r(\mathbf{e})}{s(\mathbf{e})},$$

can be obtained by the change of variable

$$(\bar{t}_1(\mathbf{e}), \ldots, \bar{t}_r(\mathbf{e})) = (t_1(\mathbf{e}), \ldots, t_r(\mathbf{e}))P^{-1}.$$

The marginal distribution of $\mathbf{t}(\mathbf{e})$ is

$$\frac{A_{n-r}}{A_n} \frac{|VV'|^{1/2}}{(1 + \mathbf{t}'VV'\mathbf{t})^{n/2}}\, d\mathbf{t}.$$

This is an r-dimensional t-distribution; note that the quadratic expression in the denominator now contains cross-product terms.

8.4 Location Distributions: Structural. The structural probability element (*orthogonal* basis) for the location quantity $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)'$ is

$$\frac{A_{n-r}}{A_n} \frac{|WW'|^{1/2} s^{-r}(\mathbf{y})}{\left(1 + (\boldsymbol{\alpha} - \mathbf{a}(\mathbf{y}))' \dfrac{WW'}{s^2(\mathbf{y})} (\boldsymbol{\alpha} - \mathbf{a}(\mathbf{y}))\right)^{n/2}}\, d\boldsymbol{\alpha};$$

this is a relocated and rescaled multivariate t-distribution.

The structural probability element (*original* basis) for the location quantity $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$ is

$$\frac{A_{n-r}}{A_n} \frac{|VV'|^{1/2} s^{-r}(\mathbf{y})}{\left(1 + (\boldsymbol{\beta} - \mathbf{b}(\mathbf{y}))' \dfrac{VV'}{s^2(\mathbf{y})} (\boldsymbol{\beta} - \mathbf{b}(\mathbf{y}))\right)^{n/2}}\, d\boldsymbol{\beta};$$

this is a relocated and rescaled t-distribution (with cross-product terms).

8.5 The Example. Consider the example with three structural vectors (Sections 1 and 6) and suppose that the error variable is standard normal. The regression model for the example is

$$e_1 = z_1 \ldots, e_{12} = z_{12},$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 7 & 12 & 8 & 15 & 13 & 14 & 17 & 14 & 14 & 17 & 14 & 17 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta_1 & \beta_2 & \beta_3 & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} \end{bmatrix},$$

where $z_1, \ldots, z_{12}$ designate standard normal variables. The arithmetical calculations from Section 6 are

$$\begin{array}{cccc|cccc}
12 & 9 & 9 & 162 & 1 & 0 & 0 & 0 \\
9 & 9 & 9 & 135 & 0 & 1 & 0 & 0 \\
9 & 9 & 15 & 141 & 0 & 0 & 1 & 0 \\
162 & 135 & 141 & 2302 & 0 & 0 & 0 & 1
\end{array}$$

$$\begin{array}{cccc|cccc}
1 & \frac{3}{4} & \frac{3}{4} & 13\frac{1}{2} & \frac{1}{12} & 0 & 0 & 0 \\
0 & 2\frac{1}{4} & 2\frac{1}{4} & 13\frac{1}{2} & -\frac{3}{4} & 1 & 0 & 0 \\
0 & 2\frac{1}{4} & 8\frac{1}{4} & 19\frac{1}{2} & -\frac{3}{4} & 0 & 1 & 0 \\
0 & 13\frac{1}{2} & 19\frac{1}{2} & 115 & -13\frac{1}{2} & 0 & 0 & 1
\end{array}$$

$$\begin{array}{cccc|cccc}
1 & 0 & 0 & 9 & \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\
0 & 1 & 1 & 6 & -\frac{1}{3} & \frac{4}{9} & 0 & 0 \\
0 & 0 & 6 & 6 & 0 & -1 & 1 & 0 \\
0 & 0 & 6 & 34 & -9 & -6 & 0 & 1
\end{array}$$

$$\begin{array}{cccc|cccc}
1 & 0 & 0 & 9 & \frac{1}{3} & -\frac{1}{3} & 0 & 0 \\
0 & 1 & 0 & 5 & -\frac{1}{3} & \frac{11}{18} & -\frac{1}{6} & 0 \\
0 & 0 & 1 & 1 & 0 & -\frac{1}{6} & \frac{1}{6} & 0 \\
0 & 0 & 0 & 28 & -9 & -5 & -1 & 1
\end{array}$$

The new quantities relative to the orthogonal basis are given by

$$(\alpha_1, \alpha_2, \alpha_3) = (\beta_1, \beta_2, \beta_3)\begin{pmatrix} 1 & 0 & 0 \\ -\frac{3}{4} & 1 & 0 \\ 0 & -1 & 0 \end{pmatrix}^{-1}$$

$$= (\beta_1, \beta_2, \beta_3)\begin{pmatrix} 1 & 0 & 0 \\ \frac{3}{4} & 1 & 0 \\ \frac{3}{4} & 1 & 1 \end{pmatrix}.$$

The reduced model relative to the orthogonal basis is

$$a_1(e) = \frac{z_1}{\sqrt{12}}, \qquad a_2(e) = \frac{z_2}{\sqrt{2\frac{1}{4}}}, \qquad a_3(e) = \frac{z_3}{\sqrt{6}}, \qquad s(e) = \chi_9,$$

$$\begin{aligned}
13\tfrac{1}{2} &= \alpha_1 + \sigma a_1(e), \\
6 &= \alpha_2 + \sigma a_2(e), \\
1 &= \alpha_3 + \sigma a_3(e), \\
\sqrt{28} &= \qquad \sigma s(e),
\end{aligned}$$

where z_1, z_2, z_3 *de novo* designate standard normal variables and χ_9 is a chi variable on nine degrees of freedom. The reduced model can also be presented relative to the corresponding *orthonormal* basis

$$z_1, \qquad z_2, \qquad z_3, \qquad \chi_9,$$

$$\begin{aligned}
13\tfrac{1}{2}\sqrt{12} &= \alpha_1\sqrt{12} + \sigma z_1, \\
6\sqrt{2\tfrac{1}{4}} &= \alpha_2\sqrt{2\tfrac{1}{4}} + \sigma z_2, \\
1\sqrt{6} &= \alpha_3\sqrt{6} + \sigma z_3, \\
\sqrt{28} &= \qquad \sigma\chi_9
\end{aligned}$$

In this alternative form the adjusted regression coefficients $13\frac{1}{2}\sqrt{12}$, $6\sqrt{2\frac{1}{4}}$, $1\sqrt{6}$ measure Euclidean distance in the directions defined by the orthogonal basis and they produce directly the components in the analysis-of-variance table:

Source	Dimension	Component	Structure of Component
Mean (w_1)	1	2187	$(\alpha_1\sqrt{12} + \sigma z_1)^2$
Treatment (w_2)	1	81	$(\alpha_2\sqrt{2\frac{1}{4}} + \sigma z_2)^2$
Variable (w_3)	1	6	$(\alpha_3\sqrt{6} + \sigma z_3)^2$
Residual (d)	9	28	$(\sigma\chi_9)^2$
	12	2302	

Consider whether the variable x affects the response level. The *hypothesis* $\beta_3 = 0$ or, equivalently, the *hypothesis* $\alpha_3 = 0$ leads to the value of an error characteristic:

$$\begin{aligned}
1 &= \sigma a_3(e), \\
\sqrt{28} &= \sigma s(e),
\end{aligned}$$

$$t_3(e) = \frac{a_3(e)}{s(e)} = \frac{1}{\sqrt{28}};$$

the corresponding error variable is

$$t_3(e) = \frac{a_3(e)}{s(e)} = \frac{z_3/\sqrt{6}}{\chi_9}.$$

This can be related to an ordinary t-variable on nine degrees of freedom: The value is

$$t^* = \frac{\sqrt{6}}{1/\sqrt{9}} \frac{1}{\sqrt{28}} = 1.388,$$

and the variable is

$$t^* = \frac{z_3}{\chi_9/\sqrt{9}}.$$

The value is a reasonable value for the distribution: The observations are in accord with the hypothesis $\alpha_3 = 0$.

The hypothesis can also be assessed from the analysis-of-variance table. The *hypothesis* $\alpha_3 = 0$ leads to the value

$$\frac{6}{28}$$

for the variable.

$$\frac{(\sigma z_3)^2}{(\sigma \chi_9)^2} = \frac{z_3^2}{\chi_9^2}.$$

This can be related to an ordinary F-variable on one over nine degrees of freedom: The value is

$$\frac{6/1}{28/9} = 1.929,$$

and the variable is

$$F = \frac{z_3^2}{\chi_9^2/9}.$$

This is equivalent to the preceding test.

Suppose now that the effect of the variable x is negligible. The model then becomes a regression model with *two* structural vectors, and the effect of the *treatment* can be assessed.

The *hypothesis* $\alpha_2 = 0$ can be tested in the preceding manner, or the structural distribution for the treatment quantity α_2 can be derived. Consider the structural distribution. The structural distribution can be derived within the model that has two structural vectors. As a reasonable precaution, however, it is preferable to derive it within the larger model having three structural vectors, in case the effect of α_3 is not negligible. The structural distribution is

$$\begin{aligned}\alpha_2 &= 6 - \frac{\sqrt{28}}{s(e)} a_2(e) \\ &= 6 - \frac{\sqrt{28}}{\chi_9} \frac{z_2}{\sqrt{2\frac{1}{4}}} \\ &= 6 - \frac{\sqrt{28}}{\sqrt{2\frac{1}{4}}\sqrt{9}} t^* \\ &= 6 - 1.17t^*.\end{aligned}$$

The distribution has the form of a t-distribution on nine degrees-of-freedom, but relocated at $\alpha_2 = 6$ and rescaled by the factor 1.17.

THE PROGRESSION MODEL†

9 THE MODEL

Consider a stable system with a sequence of response variables $y_1, \ldots, y_p$. Suppose the internal error of the system produces a sequence of errors $e_1, \ldots, e_p$ with a known distribution $f(e_1, \ldots, e_p)$ on R^p. Suppose also that the sequence of errors *progressively* affects the sequence of response variables. As system characteristics for the first response component let μ_1 be the general level and $\sigma_{(1)}$ be the response scaling of the first error component; and for the second response let μ_2 be the general level, $\sigma_{(2)}$ be the response scaling of the second error component, and τ_{21} be the response multiple of the preceding error component; and for the pth response let μ_p be the general level, $\sigma_{(p)}$ be the response scaling of the pth error component, and $\tau_{p1}, \ldots, \tau_{p\,p-1}$ be the multiples of the preceding error components. A realized sequence of errors and the corresponding sequence of response values are then connected by the equations

$$\begin{aligned}y_1 &= \mu_1 + \sigma_{(1)}e_1, \\ y_2 &= \mu_2 + \tau_{21}e_1 + \sigma_{(2)}e_2 \\ &\cdot \\ &\cdot \\ y_p &= \mu_p + \tau_{p1}e_1 + \cdots + \tau_{p\,p-1}e_{p-1} + \sigma_{(p)}e_p,\end{aligned}$$

† The remainder of this chapter may be omitted for a first reading: the methods of the regression model are used to construct a multiple response model, the progression model; the progression model has a limited range of applications but its notation and results are needed for the more useful multivariate model in Chapter Five. The material here can be read as preliminary material for Sections 4, 5, 6 in Chapter Five.

or by the matrix equation

$$\begin{bmatrix} 1 \\ y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_p \end{bmatrix} = \begin{bmatrix} 1 & & & & 0 \\ \mu_1 & \sigma_{(1)} & & & \\ \mu_2 & \tau_{21} & \sigma_{(2)} & & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ \cdot & & & & \\ \mu_p & \tau_{p1} & \cdots & \tau_{p\,p-1} & \sigma_{(p)} \end{bmatrix} \begin{bmatrix} 1 \\ e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_p \end{bmatrix}.$$

Now suppose there have been n performances of the system with observations $\mathbf{y}_1 = (y_{11}, \ldots, y_{1n})'$ for the first response, $\mathbf{y}_2 = (y_{21}, \ldots, y_{2n})'$ for the second response, ..., and $\mathbf{y}_p = (y_{p1}, \ldots, y_{pn})'$ for the pth response. The system and the n performances can then be described by the

Progression Model

$$\prod_1^n f(e_{1i}, \ldots, e_{pi}) \prod_1^n de_{1i} \cdots \prod_1^n de_{pi},$$

$$\begin{bmatrix} \mathbf{1}' \\ \mathbf{y}_1' \\ \mathbf{y}_2' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_p' \end{bmatrix} = \begin{bmatrix} 1 & & & & 0 \\ \mu_1 & \sigma_{(1)} & & & \\ \mu_2 & \tau_{21} & \sigma_{(2)} & & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot \\ \cdot & & & & \\ \mu_p & \tau_{p1} & \cdots & \tau_{p\,p-1} & \sigma_{(p)} \end{bmatrix} \begin{bmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \mathbf{e}_2' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e}_p' \end{bmatrix}.$$

The model has two parts: an error distribution with $\mathbf{e}_1, \ldots, \mathbf{e}_p$ as variables; and a *structural equation* in which realized errors $\mathbf{e}_1, \ldots, \mathbf{e}_p$ determine the relation between the unknown system characteristics and the known response observations.

The notation can be made more flexible by letting

$$Y = \begin{bmatrix} \mathbf{1}' \\ \mathbf{y}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_p' \end{bmatrix} = \left[\begin{array}{ccc} 1 & \cdots & 1 \\ \hline y_{11} & \cdots & y_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{p1} & \cdots & y_{pn} \end{array}\right] = \left[\begin{array}{c} \mathbf{1}' \\ \hline \underline{Y} \\ \\ \\ \\ \end{array}\right]$$

designate the sequence of response vectors with appended $\mathbf{1}$-vector, by letting

$$E = \left[\begin{array}{c} \mathbf{1}' \\ \hline \mathbf{e}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e}_p' \end{array}\right] = \left[\begin{array}{ccc} 1 & \cdots & 1 \\ \hline e_{11} & \cdots & e_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ e_{p1} & \cdots & e_{pn} \end{array}\right] = \left[\begin{array}{c} \mathbf{1}' \\ \hline \underline{E} \\ \\ \\ \\ \end{array}\right]$$

designate the sequence of error vectors with appended $\mathbf{1}$-vector, by letting

$$\theta = \left[\begin{array}{c|ccccc} 1 & 0 & \cdots & & & 0 \\ \hline \mu_1 & \sigma_{(1)} & & & & \\ \mu_2 & \tau_{21} & \sigma_{(2)} & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ \cdot & & & & & \cdot \\ \mu_p & \tau_{p1} & \cdots & & \tau_{p\,p-1} & \sigma_{(p)} \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline \mu & \mathcal{T} \end{array}\right]$$

designate the quantity describing the system characteristics, and by letting

$$f(E)\,dE = \prod_1^n f(e_{1i}, \ldots, e_{pi}) \prod_1^n de_{1i} \cdots \prod_1^n de_{pi}$$

designate the error distribution. The progression model can then be written

$$f(E)\,dE,$$

$$Y = \theta E.$$

The transformation θ is an element of the *location-progression group*

$$G = \left\{ g = \left[\begin{array}{c|ccccc} 1 & 0 & \cdots & & & 0 \\ \hline a_1 & c_1 & & & & 0 \\ a_2 & k_{21} & c_2 & & & \\ \cdot & \cdot & & \cdot & & \\ \cdot & \cdot & & & \cdot & \\ \cdot & \cdot & & & & \cdot \\ a_p & k_{p1} & \cdots & & k_{p\,p-1} & c_p \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline \mathbf{a} & K \end{array}\right] : \begin{array}{c} -\infty < a < \infty \\ -\infty < k < \infty \\ 0 < c < \infty \end{array} \right\},$$

with group properties

$$\begin{bmatrix} 1 & 0 \\ \mathbf{a}_1 & K_1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \mathbf{a}_2 & K_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \mathbf{a}_1 + K_1\mathbf{a}_2 & K_1K_2 \end{bmatrix},$$

$$i = \begin{bmatrix} 1 & 0 \\ 0 & I \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ \mathbf{a} & K \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -K^{-1}\mathbf{a} & K^{-1} \end{bmatrix}.$$

The component matrices K form the *progression group* G_2 under matrix multiplication: The product of two positive lower triangular matrices is positive lower triangular; the inverse of a positive lower triangular matrix

$$K^{-1} = \begin{bmatrix} k^{11} & & & 0 \\ k^{21} & k^{22} & & \\ \cdot & \cdot & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ k^{p1} & \cdots & & k^{pp} \end{bmatrix}$$

is positive lower triangular (the elements of the inverse can be calculated successively from top row to bottom row, from right to left in each row). The progression model can then be written in the alternative location-scale form

$$f(\underline{E})\,d\underline{E}$$
$$\underline{Y} = [\boldsymbol{\mu}, \mathcal{T}]\underline{E}.$$

The notation $[\mathbf{a}, K]$ designates the general location-scale transformation (Problem 27, Chapter One):

$$[\mathbf{a}, K]\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_p \end{bmatrix} = \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} + K\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_p \end{bmatrix},$$

$$[\mathbf{a}, K]\underline{Y} = \mathbf{a}\mathbf{l}' + K\underline{Y};$$

note that

$$\mathbf{a}\mathbf{l}' = \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix}(1, \ldots, 1) = \begin{bmatrix} a_1 & \cdots & a_1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_p & \cdots & a_p \end{bmatrix},$$

and note that the transformation applies column by column to the *matrix* $\underline{Y}$.

The orbit of a point Y is a set $GY = \{gY\colon g \in G\}$ in R^{pn}. The orbit can be examined alternatively in R^n by considering a *point* Y in R^{pn} as a *sequence* $\mathbf{y}_1, \ldots, \mathbf{y}_p$ of p points in R^n. A transformation g has the following effect:

$$\mathbf{y}_1 \to a_1\mathbf{1} + c_1\mathbf{y}_1,$$
$$\mathbf{y}_2 \to a_2\mathbf{1} + k_{21}\mathbf{y}_1 + c_2\mathbf{y}_2,$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$\mathbf{y}_p \to a_p\mathbf{1} + k_{p1}\mathbf{y}_1 + \cdots + k_{p\,p-1}\mathbf{y}_{p-1} + c_p\mathbf{y}_p.$$

The effect of the group on $\mathbf{y}_1$ is that of the regression-scale group with structural vector $\mathbf{1}$; the effect on $\mathbf{y}_2$ is that of the regression-scale group with structural vectors $\mathbf{1}, \mathbf{y}_1$; the effect on $\mathbf{y}_p$ is that of the regression-scale group with structural vectors $\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_{p-1}$. Suppose now that $n \geq p + 1$, and $\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p$ are linearly independent. The orbit of the first point in the sequence is $L^+(\mathbf{1}; \mathbf{y}_1)$, of the second point is $L^+(\mathbf{1}, \mathbf{y}_1; \mathbf{y}_2), \ldots,$ and of the pth point is $L^+(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_{p-1}; \mathbf{y}_p)$:

$$G \cdot \mathbf{y}_1 = L^+(\mathbf{1}; \mathbf{y}_1)$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$G \cdot \mathbf{y}_p = L^+(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_{p-1}; \mathbf{y}_p).$$

The vectors $\mathbf{y}_1, \ldots, \mathbf{y}_p$ added successively to the 1-vector generate $2, \ldots, p + 1$ dimensional spaces; the orbits for the successive elements in the sequence $\mathbf{y}_1, \ldots, \mathbf{y}_p$ are the positive halves of these spaces (see Figure 12).

Suppose a transformation g carries $\mathbf{y}_1, \ldots, \mathbf{y}_p$ into $\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_p$. The assumption that $\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p$ are linearly independent ensures that g is the only such transformation. It follows that G is unitary (linearly dependent sequences excluded); and it follows then that *the progression model is a structural model.*

10 A TRANSFORMATION VARIABLE

The transformation variable for the regression-scale group can be used to construct a transformation variable for the location-progression group.

For the first response $\mathbf{y}_1$ let $m_1(Y)$ be the regression coefficient of $\mathbf{y}_1$ on $\mathbf{1}$, let $s_{(1)}(Y)$ be the residual length, and let $\mathbf{d}_1(Y)$ be the unit residual vector:

$$\mathbf{y}_1 = m_1(Y)\mathbf{1} + s_{(1)}(Y)\,\mathbf{d}_1(Y).$$

The unit residual vector $\mathbf{d}_1(Y)$ is a fixed vector in $L^+(\mathbf{1}; \mathbf{y}_1)$.

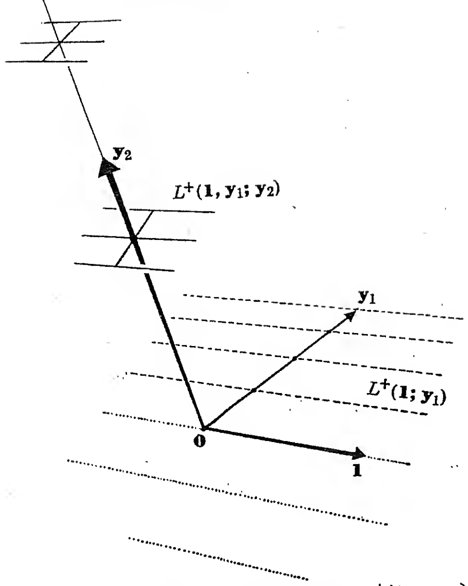


Figure 12 The orbit of $\mathbf{y}_1$ is $L^+(\mathbf{1}; \mathbf{y}_1)$; the orbit of $\mathbf{y}_2$ is $L^+(\mathbf{1}, \mathbf{y}_1; \mathbf{y}_2), \ldots$.

For the second response $\mathbf{y}_2$ let $m_2(Y)$, $t_{21}(Y)$ be the regression coefficients of $\mathbf{y}_2$ on $\mathbf{1}$, $\mathbf{d}_1(Y)$, let $s_{(2)}(Y)$ be the residual length, and let $\mathbf{d}_2(Y)$ be the unit residual vector:

$$\mathbf{y}_2 = m_2(Y)\mathbf{1} + t_{21}(Y)\mathbf{d}_1(Y) + s_{(2)}(Y)\mathbf{d}_2(Y).$$

The unit residual vector $\mathbf{d}_2(Y)$ is a fixed vector in $L^+(\mathbf{1}, \mathbf{y}_1; \mathbf{y}_2)$.

For the pth response let $m_p(Y)$, $t_{p1}(Y), \ldots, t_{p\,p-1}(Y)$ be the regression coefficients of $\mathbf{y}_p$ on $\mathbf{1}$, $\mathbf{d}_1(Y), \ldots, \mathbf{d}_{p-1}(Y)$, let $s_{(p)}(Y)$ be the residual length, and let $\mathbf{d}_p(Y)$ be the unit residual vector:

$$\mathbf{y}_p = m_p(Y)\mathbf{1} + t_{p1}(Y)\mathbf{d}_1(Y) + \cdots + t_{p\,p-1}(Y)\mathbf{d}_{p-1}(Y) + s_{(p)}(Y)\mathbf{d}_p(Y).$$

The unit residual vector $\mathbf{d}_p(Y)$ is a fixed vector in $L^+(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_{p-1}; \mathbf{y}_p)$.

The orthogonality and normality give

$$m_1(Y) = \frac{\sum_1^n y_{1i}}{n} = \bar{y}_1, \ldots, m_p(Y) = \frac{\sum_1^n y_{pi}}{n} = \bar{y}_p,$$

$$t_{jj'}(Y) = \sum_{i=1}^n y_{ji}\, d_{j'i}(Y), \qquad j' < j.$$

The coefficients and lengths can be obtained by applying the computation methods of Section 5 to the inner-product matrix $Q(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p)$.

Now let

$$[Y] = \left[\begin{array}{c|cccc} 1 & 0 & \cdots & & 0 \\ \hline m_1(Y) & s_{(1)}(Y) & & & 0 \\ m_2(Y) & t_{21}(Y) & s_{(2)}(Y) & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ m_p(Y) & t_{p1}(Y) & \cdots & t_{p\,p-1}(Y) & s_{(p)}(Y) \end{array}\right]$$

$$= \left[\begin{array}{c|c} 1 & 0 \\ \hline \mathbf{m}(Y) & T(Y) \end{array}\right]$$

and

$$D(Y) = \begin{bmatrix} \mathbf{1}' \\ \hline \mathbf{d}_1'(Y) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{d}_p'(Y) \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \hline d_{11}(Y) & \cdots & d_{1n}(Y) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ d_{p1}(Y) & \cdots & d_{pn}(Y) \end{bmatrix} = \begin{bmatrix} \mathbf{1}' \\ \hline \\ \underline{D}(Y) \\ \\ \end{bmatrix}.$$

The equations in the preceding paragraphs can then be collected in a single matrix equation:

$$Y = \begin{bmatrix} \mathbf{1}' \\ \mathbf{y}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_p' \end{bmatrix} = [Y]D(Y).$$

The variable $[Y]$ is an element in the location-progression group; the point $D(Y)$ is a point on the orbit of Y, a *fixed* point. It follows that $D(Y)$ is a reference point and $[Y]$ is a transformation variable.

In the alternative notation the equation $Y = [Y]D(Y)$ becomes

$$\underline{Y} = [\mathbf{m}(Y), T(Y)]\underline{D}(Y) = \mathbf{m}(Y)\mathbf{1}' + T(Y)\underline{D}(Y);$$

note that

$$T(Y)\underline{D}(Y) = \underline{Y} - \mathbf{m}(Y)\mathbf{1}' = \begin{bmatrix} y_{11} - \bar{y}_1 & \cdots & y_{1n} - \bar{y}_1 \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ y_{p1} - \bar{y}_p & \cdots & y_{pn} - \bar{y}_p \end{bmatrix}.$$

The progression model can now be written:

$$f(E)\,dE,$$

$$[Y] = \theta[E], \qquad D(Y) = D(E).$$

The structural equation, conditional on the orbit is $[Y] = \theta[E]$; it has the form

$$\begin{aligned}
\bar{y}_1 &= \mu_1 + \sigma_{(1)}\bar{e}_1, \\
s_{(1)}(Y) &= \sigma_{(1)}s_{(1)}(E); \\
\bar{y}_2 &= \mu_2 + \tau_{21}\bar{e}_1 + \sigma_{(2)}\bar{e}_2, \\
t_{21}(Y) &= \tau_{21}s_{(1)}(E) + \sigma_{(2)}t_{21}(E), \\
s_{(2)}(Y) &= \sigma_{(2)}s_{(2)}(E); \\
&\cdot \\
&\cdot \\
&\cdot \\
\bar{y}_p &= \mu_p + \tau_{p1}\bar{e}_1 + \cdots + \tau_{p\,p-1}\bar{e}_{p-1} + \sigma_{(p)}\bar{e}_p, \\
t_{p1}(Y) &= \tau_{p1}s_{(1)}(E) + \cdots + \tau_{p\,p-1}t_{p-1\,1}(E) + \sigma_{(p)}t_{p1}(E), \\
&\cdot \\
&\cdot \\
&\cdot \\
s_{(p)}(Y) &= \sigma_{(p)}s_{(p)}(E).
\end{aligned}$$

In the alternative notation the progression model can be written

$$f(E)\,dE,$$

$$[\mathbf{m}(Y), T(Y)] = [\boldsymbol{\mu}, \mathscr{C}][\mathbf{m}(E), T(E)], \qquad \underline{D}(Y) = \underline{D}(E).$$

The structural equation conditional on the orbit can then be separated into two component equations:

$$\begin{aligned}
\mathbf{m}(Y) &= \boldsymbol{\mu} + \mathscr{C}\mathbf{m}(E), \\
T(Y) &= \mathscr{C}T(E).
\end{aligned}$$

11 THE REDUCED MODEL

Consider first the invariant differential on the sample space. A transformation g operates column-by-column on Y. Its effect on the ith column is

$$[\mathbf{a}, K]\begin{bmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi} \end{bmatrix} = \mathbf{a} + K\begin{bmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi} \end{bmatrix},$$

which has Jacobian $|K| = c_1 \cdots c_p$. Hence

$$J_{pn}(g, Y) = |K|^n = |g|^n = (c_1 \cdots c_p)^n,$$

$$J_{pn}(Y) = |T(Y)|^n = |[Y]|^n = (s_{(1)}(Y) \cdots s_{(p)}(Y))^n,$$

$$dm(Y) = \frac{\prod dy_{ji}}{s^n_{(1)}(Y) \cdots s^n_{(p)}(Y)} = \frac{dY}{|[Y]|^n}.$$

Now consider the invariant differentials on the group:

$$\begin{bmatrix} 1 & & & & 0 \\ \tilde{a}_1 & \tilde{c}_1 & & & \\ \tilde{a}_2 & \tilde{k}_{21} & \tilde{c}_2 & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \\ \tilde{a}_p & \tilde{k}_{p1} & \cdots & \tilde{k}_{p\,p-1} & \tilde{c}_p \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & & 0 \\ a_1 & c_1 & & & \\ a_2 & k_{21} & c_2 & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \cdot \\ a_p & k_{p1} & \cdots & k_{p\,p-1} & c_p \end{bmatrix} \begin{bmatrix} 1 & & & & 0 \\ a^*_1 & c^*_1 & & & \\ a^*_2 & -k^*_{21} & c^*_2 & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \cdot \\ a^*_p & k^*_{p1} & \cdots & k^*_{p\,p-1} & c^*_p \end{bmatrix}.$$

The left transformation operates column-by-column:

$$J = (c_1 \cdots c_p)(c_1 \cdots c_p)(c_2 \cdots c_p) \cdots (c_p) = c_1^2 \cdots c_p^{p+1},$$

$$J(g) = c_1^2 \cdots c_p^{p+1} = |g|_\Delta,$$

where $|a_{jj'}|_\Delta$ designates the *increasing determinant*,

$$\begin{vmatrix} a_{11} & & & 0 \\ a_{21} & a_{22} & & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}_\Delta = a_{11}^1 a_{22}^2 \cdots a_{kk}^k;$$

$$d\mu(g) = \frac{\prod da_j \prod dk_{jj'} \prod dc_j}{c_1^2 \cdots c_p^{p+1}} = \frac{dg}{|g|_\Delta}.$$

The right transformation operates row-by-row:

$$J^* = (1c_1^*)(1c_1^* c_2^*) \cdots (1c_1^* \cdots c_p^*)$$

$$J^*(g) = c_1^p c_2^{p-1} \cdots c_p^1 = |g|_\nabla,$$

where $|a_{jj'}|_\nabla$ designates the *decreasing* determinant,

$$\begin{vmatrix} a_{11} & & & 0 \\ a_{21} & a_{22} & & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}_\nabla = a_{11}^k a_{22}^{k-1} \cdots a_{kk}^1;$$

$$d\nu(g) = \frac{\prod da_j \prod dk_{jj'} \prod dc_j}{c_1^p \cdots c_p^1} = \frac{dg}{|g|_\nabla}.$$

The modular function is

$$\Delta(g) = \frac{|g|_\nabla}{|g|_\Delta} = \frac{c_1^p \cdots c_p^1}{c_1^2 \cdots c_p^{p+1}}.$$

11.1 General Distributions. The conditional probability element for the error position $[E]$ given the orbit $D(E) = D$ is

$$g([E]: D)\, d[E] = k(D) f([E]D)\, \|[E]\|^n \frac{d[E]}{\|[E]\|_\Delta}$$

$$= k(D) \prod_{i=1}^n f\left(\begin{bmatrix} m_1 \\ \cdot \\ \cdot \\ \cdot \\ m_p \end{bmatrix} + T \begin{bmatrix} d_{1i} \\ \cdot \\ \cdot \\ \cdot \\ d_{pi} \end{bmatrix} \right) s_{(1)}^{n-2} \cdots s_{(p)}^{n-p-1}\, d[E].$$

The structural probability element for θ given Y is

$$g^*(\theta : Y)\, d\theta = k(D(Y)) f(\theta^{-1}Y) \frac{\|[Y]\|^n}{|\theta|^n} \frac{\|[Y]\|_\nabla}{\|[Y]\|_\Delta} d\nu(\theta)$$

$$= k(D(Y)) \prod_{i=1}^n f\left([\boldsymbol{\mu}, \mathcal{C}]^{-1} \begin{bmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi} \end{bmatrix} \right) \frac{\|[Y]\|^n}{|\theta|^n} \frac{\|[Y]\|_\nabla}{\|[Y]\|_\Delta} d\nu(\theta)$$

$$= k(D(Y)) \prod_{i=1}^n f\left(\mathcal{C}^{-1} \begin{bmatrix} y_{1i} - \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{pi} - \mu_p \end{bmatrix} \right) \left(\frac{s_{(1)}(Y) \cdots s_{(p)}(Y)}{\sigma_{(1)} \cdots \sigma_{(p)}} \right)^n$$

$$\cdot \frac{s_{(1)}^p(Y) \cdots s_{(p)}^1(Y)}{s_{(1)}^2(Y) \cdots s_{(p)}^{p+1}(Y)} \cdot \frac{d\boldsymbol{\mu}\, d\mathcal{C}}{\sigma_{(1)}^p \cdots \sigma_{(p)}^1}.$$

11.2 The structural equation for the progression model,

$$\mathbf{m}(Y) = \boldsymbol{\mu} + \mathcal{C}\mathbf{m}(E),$$

$$T(Y) = \mathcal{C}T(E),$$

can be separated into a part concerning μ and a part concerning $\mathcal{C}$:

$$T^{-1}(Y)(\mathbf{m}(Y) - \boldsymbol{\mu}) = T^{-1}(E)\mathbf{m}(E) = \mathbf{t}(E),$$

$$\mathcal{C}^{-1}T(Y) = T(E).$$

In a related manner the location-progression group has a *location subgroup*

$$L = \{[\mathbf{a}, I]: \quad \mathbf{a} \in R^p\}$$

and a *scale subgroup*

$$S = \{[0, K]: \quad K \in G_2\};$$

note that S and G_2 are the same group but differ in designation and in the spaces to which they can be applied by matrix multiplication.

11.3 Location Distributions. For inference concerning $\boldsymbol{\mu}$, the marginal distribution of the error variable $\mathbf{t} = \mathbf{t}(E) = (t_1(E), \ldots, t_p(E))'$ is needed. The full error probability distribution is

$$g([E]: D)\, d[E] = k(D) f([E]D)\, \|[E]\|^n \frac{d[E]}{\|[E]\|_\Delta}$$

$$= k(D) f([E]D)\, |T|^n \frac{d\mathbf{m}\, dT}{|T|\, |T|_\Delta};$$

the transformation $\mathbf{m} = T\mathbf{t}$ from $\mathbf{t}$ to $\mathbf{m}$ for fixed T has Jacobian $|T|$; the marginal distribution of $\mathbf{t}$ is obtained by substituting for $\mathbf{m}$ and integrating out T:

$$g_L(\mathbf{t}: D)\, d\mathbf{t} = k(D)\int_T f([E]D)\, |T|^n \frac{dT}{|T|_\Delta} \cdot d\mathbf{t}$$

$$= k(D)\int_T \prod_1^n f\left(T\begin{bmatrix} t_1 + d_{1i} \\ \cdot \\ \cdot \\ \cdot \\ t_p + d_{pi}\end{bmatrix}\right)|T|^n \frac{dT}{|T|_\Delta} \cdot d\mathbf{t}.$$

The error variable $\mathbf{t} = \mathbf{t}(E)$ indexes the orbits (right cosets) of the scale group S on G^*:

$$S[\mathbf{m}, T] = S[0, T][\mathbf{t}, I] = S[\mathbf{t}, I].$$

11.4 Scale Distributions. For inference concerning $\mathscr{T}$ the marginal distribution of the error variable $T = T(E)$ is needed. The marginal distribution is obtained from the full error distribution by integrating out $\mathbf{m}$:

$$g_S(T: D)\, dT = k(D)\int_{\mathbf{m}} f([E]D)\, d\mathbf{m} \cdot |T|^n \frac{dT}{|T|\,|T|_\Delta}$$

$$= k(D)\cdot\int_{\mathbf{m}} \prod_1^n f\left(\begin{bmatrix} m_1 \\ \cdot \\ \cdot \\ \cdot \\ m_p\end{bmatrix} + T\begin{bmatrix} d_{1i} \\ \cdot \\ \cdot \\ \cdot \\ d_{pi}\end{bmatrix}\right) d\mathbf{m} \cdot s_{(1)}^{n-2} \cdots s_{(p)}^{n-p-1}\, dT.$$

The error variable $T = T(E)$ indexes the orbits (right cosets) of the location group L on G^*:

$$L[\mathbf{m}, T] = L[\mathbf{m}, I][0, T] = L[0, T].$$

12 WITH NORMAL ERROR

Consider the progression model with standard normal error variables:

$$f(E)\, dE = (2\pi)^{-np/2} \exp\{-\tfrac{1}{2}\textstyle\sum e_{ji}^2\} \prod_{i=1}^n \prod_{j=1}^p de_{ji},$$

$$[Y] = \theta[E], \qquad D(Y) = D(E).$$

The sum of squares in the exponent of the normal density function can be expressed in matrix notation:

$$\sum_{i,j} e_{ji}^2 = \sum_{j=1}^p |\mathbf{e}_j|^2 = \sum_{j=1}^p (\mathbf{e}_j, \mathbf{e}_j)$$
$$= \operatorname{tr} \underline{EE'}$$
$$= \operatorname{tr} EE' - n.$$

The notation $\operatorname{tr} S$ designates the trace of a square matrix,

$$\operatorname{tr}\begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ s_{p1} & \cdots & s_{pp}\end{bmatrix} = \sum_{j=1}^p s_{jj}.$$

Note that $\operatorname{tr} ABC = \operatorname{tr} BCA = \operatorname{tr} CAB$,—provided the matrix operations are permissible.

The sum of squares can be expressed in terms of the position variable:

$$\sum_{i,j} e_{ji}^2 = \operatorname{tr} EE' - n$$
$$= \operatorname{tr} [E]D(E)D'(E)[E]' - n$$
$$= \operatorname{tr} [E]\begin{bmatrix} n & & & & 0 \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & 1\end{bmatrix}[E]' - n$$
$$= \operatorname{tr} \llbracket E\rrbracket\llbracket E\rrbracket' - n,$$

where

$$\llbracket E\rrbracket = [E]\begin{bmatrix} \sqrt{n} & & & & 0 \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & 1\end{bmatrix} = \begin{bmatrix} \sqrt{n} & 0 \\ \sqrt{n}\,\mathbf{m}(E) & T(E)\end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{n} & 0 & & \cdots & 0 \\ \sqrt{n}\,\bar{e}_1 & s_{(1)}(E) & & & 0 \\ \sqrt{n}\,\bar{e}_2 & t_{21}(E) & s_{(2)}(E) & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \sqrt{n}\,\bar{e}_p & t_{p1}(E) & \cdots & t_{p\,p-1}(E) & s_{(p)}(E)\end{bmatrix}$$

The row vectors in $[E]$ describe the row vectors in E, but relative to basis vectors given by the rows of $D(E)$ (compare with the concluding paragraph of Section 6); the rows of $D(E)$ form an orthonormal set, *except* for the

first row, which has length $\sqrt{n}$; the *adjusted position matrix* $[E]$ describes the row vectors of E relative to an orthonormal set; thus the inner-product matrices are equal, $EE' = [E][E]'$. The trace of EE' gives the sum of squares of elements of E; hence

$$\begin{aligned}
\sum e_{ji}^2 &= \operatorname{tr} [E][E]' - n \\
&= n \operatorname{tr} \mathbf{m}(E)\mathbf{m}'(E) + \operatorname{tr} T(E)T'(E) \\
&= \left(\sum n\bar{e}_j^2\right) + \left(\sum t_{jj'}^2(E) + \sum s_{(j)}^2(E)\right).
\end{aligned}$$

12.1 General Distribution: Error. The error probability distribution is

$$\begin{aligned}
g([E]:D)\, d[E] &= k(D)(2\pi)^{-np/2} \exp\{-\tfrac{1}{2}(\operatorname{tr} [E][E]' - n)\}\, \|[E]\|^n \frac{d[E]}{\|[E]\|_\Delta} \\
&= k'(D) \exp\{-\tfrac{1}{2}(\sum n\bar{e}_j^2 + \sum t_{jj'}^2 + \sum s_{(j)}^2)\} s_{(1)}^{n-2} \cdots s_{(p)}^{n-p-1}\, d[E] \\
&= \frac{A_{n-1} \cdots A_{n-p}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2} \sum n\bar{e}_j^2 - \tfrac{1}{2} \sum t_{jj'}^2 - \tfrac{1}{2} \sum s_{(j)}^2\} \\
&\quad \cdot s_{(1)}^{n-2} \cdots s_{(p)}^{n-p-1} \prod d\sqrt{n}\, \bar{e}_j \prod dt_{jj'} \prod ds_{(j)}.
\end{aligned}$$

The error components $\sqrt{n}\, e_{j'}$, $t_{jj'}$ are independent standard normal variables; the error components $s_{(1)}, \ldots, s_{(p)}$ are independent chi-variables on $n-1, \ldots, n-p$ degrees of freedom; the error distribution does not depend on the orbit as given by D.

The error distribution can be described by the equation

$$[E] = \begin{bmatrix}
\sqrt{n} & 0 & \cdots & & 0 \\
\sqrt{n}\,\bar{e}_1 & s_{(1)} & & & 0 \\
\sqrt{n}\,\bar{e}_2 & t_{21} & s_{(2)} & & \\
\cdot & \cdot & \cdot & & \\
\cdot & \cdot & & \cdot & \\
\cdot & \cdot & & & \\
\sqrt{n}\,\bar{e}_p & t_{p1} & & t_{p\,p-1} & s_{(p)}
\end{bmatrix}
= \begin{bmatrix}
\sqrt{n} & 0 & \cdots & & 0 \\
z_1 & \chi_{n-1} & & & 0 \\
z_2 & z_{21} & \chi_{n-2} & & \\
\cdot & \cdot & & \cdot & \\
\cdot & \cdot & & \cdot & \\
\cdot & \cdot & & & \cdot \\
z_p & z_{p1} & & z_{p\,p-1} & \chi_{n-p}
\end{bmatrix},$$

where the z's are independent standard normal variables and the χ's are chi-variables with degrees of freedom as subscribed.

12.2 General Distribution: Structural. The structural distribution for the quantity θ can be obtained from the general expression in the preceding section. The matrix form for the sum of squares in the exponent and the substitution $E = \theta^{-1} Y$ give

$$\begin{aligned}
g^*(\theta:Y)\, d\theta &= \frac{A_{n-1} \cdots A_{n-p}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2}(\operatorname{tr} \theta^{-1} YY'\theta'^{-1} - n)\} \\
&\quad \cdot \left(\frac{s_{(1)}(Y) \cdots s_{(p)}(Y)}{\sigma_{(1)} \cdots \sigma_{(p)}}\right)^n \frac{s_{(1)}^p(Y) \cdots s_{(p)}^1(Y)}{s_{(1)}^2(Y) \cdots s_{(p)}^{p+1}(Y)} \frac{n^{p/2}\, d\boldsymbol{\mu}\, d\mathcal{T}}{\sigma_{(1)}^p \cdots \sigma_{(p)}^1}.
\end{aligned}$$

The location and scale components in the exponential can be expressed in terms of the quantities $\boldsymbol{\mu}$, $\mathcal{T}$. The formulas

$$\begin{aligned}
\mathbf{m}(Y) &= \boldsymbol{\mu} + \mathcal{T}\mathbf{m}, \\
T(Y) &= \quad\;\; \mathcal{T}T
\end{aligned}$$

can be used in the expression for the sum of squares of the error variables:

$$\begin{aligned}
\sum e_{ji}^2 &= \operatorname{tr} EE' - n \\
&= n \operatorname{tr} \mathbf{m}\mathbf{m}' + \operatorname{tr} TT' \\
&= n \operatorname{tr} \mathcal{T}^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu})(\mathbf{m}(Y) - \boldsymbol{\mu})'\mathcal{T}'^{-1} + \operatorname{tr} \mathcal{T}^{-1}T(Y)T'(Y)\mathcal{T}'^{-1} \\
&= n(\mathbf{m}(Y) - \boldsymbol{\mu})'(\mathcal{T}\mathcal{T}')^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu}) + \operatorname{tr} (\mathcal{T}\mathcal{T}')^{-1}(T(Y)T'(Y)) \\
&= n(\mathbf{m}(Y) - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu}) + \operatorname{tr} \Sigma^{-1}S(Y).
\end{aligned}$$

The inner-product matrix $\mathcal{T}\mathcal{T}'$ is designated

$$\Sigma = \mathcal{T}\mathcal{T}',$$

and the inner-product matrix $T(Y)T'(Y)$ is designated

$$\begin{aligned}
S(Y) &= T(Y)T'(Y) = T(Y)\underline{D}(Y)\underline{D}'(Y)T'(Y) \\
&= \begin{bmatrix}
y_{11} - \bar{y}_1 & \cdots & y_{1n} - \bar{y}_1 \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\cdot & & \cdot \\
y_{p1} - \bar{y}_p & \cdots & y_{pn} - \bar{y}_p
\end{bmatrix}
\begin{bmatrix}
y_{11} - \bar{y}_1 & \cdots & y_{1n} - \bar{y}_1 \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\cdot & & \cdot \\
y_{p1} - \bar{y}_p & \cdots & y_{pn} - \bar{y}_p
\end{bmatrix}'.
\end{aligned}$$

Note that $S(Y)$ is the inner product matrix for the response deviation vectors $\mathbf{y}_j - \bar{y}_j\mathbf{1} = (y_{j1} - \bar{y}_j, \ldots, y_{jn} - \bar{y}_j)'$.

The structural distribution can now be written:

$$
\begin{aligned}
g^*(\theta : Y)\,d\theta &\\
&= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2}(\mathbf{m}(Y)-\boldsymbol{\mu})' n\,\Sigma^{-1}(\mathbf{m}(Y)-\boldsymbol{\mu}) - \tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\\
&\quad\cdot \left(\frac{s_{(1)}(Y)\cdots s_{(p)}(Y)}{\sigma_{(1)}\cdots\sigma_{(p)}}\right)^n \frac{s^p_{(1)}(Y)\cdots s^1_{(p)}(Y)}{s^2_{(1)}(Y)\cdots s^{p+1}_{(p)}(Y)} \frac{n^{p/2}\,d\boldsymbol{\mu}\,d\mathscr{C}}{\sigma^p_{(1)}\cdots\sigma^1_{(p)}}.
\end{aligned}
$$

Note that the conditional distribution of $\boldsymbol{\mu}$ given $\mathscr{C}$ is normal with mean $\mathbf{m}(Y)$ and covariance matrix $n^{-1}\Sigma$; the marginal distribution for $\mathscr{C}$ can be described in terms of normal and chi-variables by the relation

$$T(Y) = \mathscr{C}T(E):$$

$$
\mathscr{C} = T(Y)\begin{bmatrix} \chi_{n-1} & & & 0 \\ z_{21} & \chi_{n-2} & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ z_{p1} & \cdots & z_{p\,p-1} & \chi_{n-p} \end{bmatrix}^{-1}.
$$

12.3 The Error Inner-Product Distribution. The error scale matrix $T(E)$ has a distribution described by

$$
T(E) = \begin{bmatrix} s_{(1)} & & & 0 \\ t_{21} & s_{(2)} & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ t_{p1} & \cdots & t_{p\,p-1} & s_{(p)} \end{bmatrix} = \begin{bmatrix} \chi_{n-1} & & & 0 \\ z_{21} & \chi_{n-2} & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ z_{p1} & \cdots & z_{p\,p-1} & \chi_{n-p} \end{bmatrix};
$$

with probability element

$$\frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}} \exp\{-\tfrac{1}{2}\sum t^2_{jj'} - \tfrac{1}{2}\sum s^2_{(j)}\} s^{n-2}_{(1)}\cdots s^{n-p-1}_{(p)}\,dT.$$

A closely related matrix is the error inner-product matrix:

$$
\begin{aligned}
S(E) &= T(E)T'(E) = T(E)\underline{D}(E)\underline{D}'(E)T'(E)\\
&= \begin{bmatrix} e_{11}-\bar e_1 & \cdots & e_{1n}-\bar e_1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ e_{p1}-\bar e_p & \cdots & e_{pn}-\bar e_p \end{bmatrix} \begin{bmatrix} e_{11}-\bar e_1 & \cdots & e_{1n}-\bar e_1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ e_{p1}-\bar e_p & \cdots & e_{pn}-\bar e_p \end{bmatrix}'.
\end{aligned}
$$

The definitions in Section 10 identify $s_{(1)}$ as the length of the first deviation vector, t_{21} as the regression coefficient of the second vector on the unit first vector and $s_{(2)}$ as the residual length, . . . , $t_{p1}, \ldots, t_{p\,p-1}$ as the regression coefficients of the pth vector on the unit orthogonal vectors derived successively from the first $p-1$ vectors and $s_{(p)}$ as the residual length. Thus $T(E)$ can be derived from $S(E)$ as the unique *positive lower triangular square root* (compare the end of Section 6). It follows that there is a one-to-one correspondence between the matrices $T(E)$ and $S(E)$.

Consider the probability distribution for the inner product matrix:

$$
S(E) = \begin{bmatrix} s_{11}(E) & \cdots & s_{1p}(E) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ s_{p1}(E) & \cdots & s_{pp}(E) \end{bmatrix},
$$

a symmetric matrix with $s_{jj'} = s_{j'j}$. The transformation from T to S,

$$
\begin{bmatrix} s_{11} & * & \cdots & * \\ s_{21} & s_{22} & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & & * \\ \cdot & & \cdot & \\ s_{p1} & \cdots & & s_{pp} \end{bmatrix} = \begin{bmatrix} t_{11} & & & 0 \\ t_{21} & t_{22} & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ t_{p1} & \cdots & & t_{pp} \end{bmatrix} \begin{bmatrix} t_{11} & t_{21} & \cdots & t_{p1} \\ & t_{22} & & \cdot \\ & & \cdot & \cdot \\ & & & \cdot \\ 0 & & & t_{pp} \end{bmatrix},
$$

has *lower-triangular* form when the variables are taken in the order s_{11}; $s_{21}, s_{22}; \ldots; s_{p1}, \ldots, s_{pp}$:

$$
\begin{array}{lll|l}
\underline{s_{11}} = t^2_{\underline{11}} & & & 2t_{11} \\
\underline{s_{21}} = & t_{11}t_{\underline{21}} & & t_{11} \\
\underline{s_{22}} = & t^2_{21} + t^2_{\underline{22}} & & 2t_{22} \\
\cdot & & & \cdot \\
\cdot & & & \\
\cdot & & & \\
\underline{s_{p1}} & t_{11}t_{\underline{p1}} & & t_{11} \\
\underline{s_{p2}} = & t_{21}t_{p1} + t_{22}t_{\underline{p2}} & & t_{22} \\
\cdot & & & \cdot \\
\cdot & & & \cdot \\
\cdot & & & \cdot \\
\underline{s_{pp}} = & t^2_{p1} + t^2_{p2} + \cdots + t^2_{\underline{pp}} & & 2t_{pp}.
\end{array}
$$

The Jacobian matrix is lower-triangular, and the Jacobian determinant is then the product of the diagonal elements (the diagonal elements are recorded in the right margin):

$$\left|\frac{\partial S}{\partial T}\right| = 2t_{11}(t_{11}2t_{22})\cdots(t_{11}t_{22}\cdots 2t_{pp})$$
$$= 2^p\,|T|_\nabla.$$

The probability element for S is obtained by substitution:

$$\frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr} TT'\}\,|T|^{n-1}\frac{dT}{|T|_\Delta}$$
$$= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr} S\}\,|S|^{(n-1)/2}\frac{dS}{|T|_\Delta 2^p |T|_\nabla}$$
$$= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr} S\}\,|S|^{(n-1)/2}\frac{dS}{2^p|S|^{(p+1)/2}}$$
$$= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}(s_{11}+\cdots+s_{pp}\}\,|S|^{(n-1-(p+1))/2}$$
$$\cdot\frac{ds_{11}(ds_{21}\,ds_{22})\cdots(ds_{p1}\cdots ds_{pp})}{2^p}.$$

The density applies to all points S for which the matrix S is positive definite. The distribution of S is the *standard Wishart distribution.*

NOTES AND REFERENCES

The regression model as developed in this chapter can be used to generate a corresponding classical model $f(\mathbf{y}:\boldsymbol{\beta},\sigma)$. The classical model $f(\mathbf{y}:\boldsymbol{\beta},\sigma)$ gives the distribution of possible response vectors $\mathbf{y}$ based on a fixed value of the quantity $(\boldsymbol{\beta},\sigma)$:

$$\prod_{i=1}^n f(e_i)\prod_{i=1}^n de_i = \prod_{i=1}^n f(e_i)s_e^n\frac{d\mathbf{e}}{s_e^n}$$
$$= \prod_{i=1}^n f\left(\frac{y_i-\sum\beta_u v_{ui}}{\sigma}\right)\left(\frac{s_y}{\sigma}\right)^n\frac{d\mathbf{y}}{s_y^n}$$
$$= \prod_{i=1}^n f\left(\frac{y_i-\sum\beta_u v_{ui}}{\sigma}\right)\frac{1}{\sigma^n}\,d\mathbf{y} = f(\mathbf{y}:\boldsymbol{\beta},\sigma)\,d\mathbf{y}.$$

The classical regression model *with normal error* has an extensive literature. Its essential form appears in the work of Legendre, Gauss, and Laplace and is tied closely to the classical methods of least squares. Some current books on the topic are Plackett (1960), Scheffé (1959), and Williams (1959).

The classical model with other error forms has received little attention. Classical theory does not produce a set of natural variables to be in correspondence with the *parameters* $\boldsymbol{\beta}$, σ. And the use of the regression variables of normal and least-squares theory leads to the intractable problem of finding the marginal distribution of these variables. The little attention that has been accorded the nonnormal model has been concerned with examining the methods of normal theory as used with an error form that departs modestly from normality.

The structural distribution for $(\boldsymbol{\beta},\sigma)$ was developed in another framework as a distribution for $(\boldsymbol{\beta},\sigma)$: Fraser (1961), Verhagen (1961). The development here follows closely that in Fraser (1967).

The progression model also leads to a corresponding classical model. For single values on each response variable, the distribution describing possible $(y_1,\ldots,y_p)$ for given $\boldsymbol{\mu}$, $\mathcal{C}$ is

$$f(e_1,\ldots,e_p)\prod de_j = f\left(\mathcal{C}^{-1}\begin{bmatrix}y_1-\mu_1\\ \cdot\\ \cdot\\ \cdot\\ y_p-\mu_p\end{bmatrix}\right)\frac{\prod dy_j}{|\mathcal{C}|}$$
$$= f(y_1,\ldots,y_p:\boldsymbol{\mu},\mathcal{C})\prod dy_j.$$

With standard normal error form the distribution for $(y_1,\ldots,y_p)$ is

$$f(y_1,\ldots,y_p:\boldsymbol{\mu},\mathcal{C})\prod dy_j$$
$$= \frac{1}{(2\pi)^{p/2}}\exp\left\{-\tfrac{1}{2}\operatorname{tr}\mathcal{C}^{-1}\begin{bmatrix}y_1-\mu_1\\ \cdot\\ \cdot\\ \cdot\\ y_p-\mu_p\end{bmatrix}\begin{bmatrix}y_1-\mu_1\\ \cdot\\ \cdot\\ \cdot\\ y_p-\mu_p\end{bmatrix}'\mathcal{C}'^{-1}\right\}\frac{\prod dy_j}{|\mathcal{C}|}$$
$$= \frac{|\Sigma|^{-1/2}}{(2\pi)^{p/2}}\exp\left\{-\frac{1}{2}\begin{bmatrix}y_1-\mu_1\\ \cdot\\ \cdot\\ \cdot\\ y_p-\mu_p\end{bmatrix}'\Sigma^{-1}\begin{bmatrix}y_1-\mu_1\\ \cdot\\ \cdot\\ \cdot\\ y_p-\mu_p\end{bmatrix}\right\}\prod dy_j;$$

the distribution is multivariate normal with mean $(\mu_1,\ldots,\mu_p)$ and covariance matrix $\Sigma = \mathcal{C}\mathcal{C}'$.

For n values on each response variable, the distribution describing $\underline{Y}$ is

$$f(\underline{Y}:\boldsymbol{\mu},\Sigma)\,d\underline{Y} = \frac{|\Sigma|^{-n/2}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2} \operatorname{tr}(\underline{Y}-\boldsymbol{\mu}\mathbf{1}')'\Sigma^{-1}(\underline{Y}-\boldsymbol{\mu}\mathbf{1}')\}\,d\underline{Y}.$$

The marginal distribution of $\mathbf{m}(Y) = (\bar{y}_1, \ldots, \bar{y}_p)'$ can be obtained by the change of variable $\mathbf{m}(Y) = \boldsymbol{\mu} + \mathscr{C}\mathbf{m}$ from the conditional (also marginal) distribution of $\mathbf{m} = \mathbf{m}(E)$:

$$f(\mathbf{m}(Y):\boldsymbol{\mu},\Sigma)\,d\mathbf{m}(Y) = \frac{n^{p/2}|\Sigma|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}\begin{pmatrix}\bar{y}_1-\mu_1\\ \cdot\\ \cdot\\ \cdot\\ \bar{y}_p-\mu_p\end{pmatrix}' n\Sigma^{-1}\begin{pmatrix}\bar{y}_1-\mu_1\\ \cdot\\ \cdot\\ \cdot\\ \bar{y}_p-\mu_p\end{pmatrix}\right\}\prod d\bar{y}_{j\cdot}\,.$$

The marginal distribution of $T(Y)$ can be obtained by the change of variable $T(Y) = \mathscr{C}T$ from the conditional (also marginal) distribution of $T = T(E)$:

$$\begin{aligned}
&\frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}TT'\}\,|T|^{n-1}\frac{dT}{|T|_\Delta}\\
&= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\mathscr{C}^{-1}T(Y)T'(Y)\mathscr{C}'^{-1}\}\frac{|T(Y)|^{n-1}}{|\mathscr{C}|^{n-1}}\frac{dT(Y)}{|T(Y)|_\Delta}\\
&= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\frac{|S(Y)|^{(n-1)/2}}{|\Sigma|^{(n-1)/2}}\frac{dT(Y)}{|T(Y)|_\Delta}\\
&= f(T(Y):\boldsymbol{\mu},\Sigma)\,dT(Y).
\end{aligned}$$

The variables $\mathbf{m}(Y)$ and $T(Y)$ are statistically independent. The distribution of the inner product matrix $S(Y)$ can be derived by the results in Section 12:

$$\begin{aligned}
&f(S(Y):\boldsymbol{\mu},\Sigma)\,dS(Y)\\
&= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\frac{|S(Y)|^{(n-1)/2}}{|\Sigma|^{(n-1)/2}}\frac{dS(Y)}{2^p\,|S(Y)|^{(p+1)/2}}\,.
\end{aligned}$$

This is the *general Wishart distribution* with covariance matrix Σ and degrees of freedom $n - 1$.

The progression model with normal error leads to the classical multivariate-normal distribution and to related distributions for the sample means and the sample covariance matrix. The internal structuring provided here by the progression model is, however, rather specialized; a more generally appropriate structural model relating to the multivariate normal is examined in Chapter Five.

Fraser, D. A. S. (1961), On fiducial inference, *Ann. Math. Statistics*, **32**, 661–676.

Fraser, D. A. S. (1967), Data transformations and the linear model, *Ann. Math. Statistics*, **38**, 1456-1465.

Mauldon, J. G. (1955), Pivotal quantities for Wishart's and related distributions and a paradox in fiducial theory, *J. Roy. Statist. Soc.*, **B17**, 79–85.

Plackett, R. L. (1960), *Regression Analysis*, Oxford University Press, London.

Scheffé, H. (1959), *The Analysis of Variance*, Wiley, New York.

Verhagen, A. M. W. (1961), The estimation of regression and error-scale parameters, when the joint distribution of the errors is of any continuous form and known apart from a scale parameter, *Biometrika*, **48**, 125–132.

Williams, E. J. (1959), *Regression Analysis*, Wiley, New York.

PROBLEMS

1. Consider a measuring instrument with error distribution $f(e)\,de$ and suppose that four measurements have been made on a quantity μ_1, three measurements on a quantity μ_2, and three measurements on a quantity μ_3:

$$\prod f(e_{ij})\prod de_{ij},$$

$$\begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1\\
y_{11} & y_{12} & y_{13} & y_{14} & y_{21} & y_{22} & y_{23} & y_{31} & y_{32} & y_{33}
\end{bmatrix}$$

$$= \begin{bmatrix}
1 & 0 & 0 & 0\\
0 & 1 & 0 & 0\\
0 & 0 & 1 & 0\\
\mu_1 & \mu_2 & \mu_3 & \sigma
\end{bmatrix}
\begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1\\
e_{11} & e_{12} & e_{13} & e_{14} & e_{21} & e_{22} & e_{23} & e_{31} & e_{32} & e_{33}
\end{bmatrix}.$$

Note that the structural vectors, designated $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, are mutually orthogonal.

(i) Show that the projection of $\mathbf{y}$ into the subspace $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is

$$\bar{y}_1\mathbf{v}_1 + \bar{y}_2\mathbf{v}_2 + \bar{y}_3\mathbf{v}_3 = (\bar{y}_1, \bar{y}_1, \bar{y}_1, \bar{y}_1, \bar{y}_2, \bar{y}_2, \bar{y}_2, \bar{y}_3, \bar{y}_3, \bar{y}_3)'$$

where $\bar{y}_1 = \sum y_{1j}/4$, $\bar{y}_2 = \sum y_{2j}/3$, $\bar{y}_3 = \sum y_{3j}/3$. This projection gives the location of $\mathbf{y}$ relative to the subspace $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$.

(ii) Check that the residual vector is

$$\begin{aligned}
s(\mathbf{y})\mathbf{d}(\mathbf{y}) = (&y_{11}-\bar{y}_1, y_{12}-\bar{y}_1, y_{13}-\bar{y}_1, y_{14}-\bar{y}_1,\\
&y_{21}-\bar{y}_2, y_{22}-\bar{y}_2, y_{23}-\bar{y}_2, y_{31}-\bar{y}_3, y_{32}-\bar{y}_3, y_{33}-\bar{y}_3)'
\end{aligned}$$

and show that the squared residual length is

$$\begin{aligned}
s^2(\mathbf{y}) &= \sum (y_{1j}-\bar{y}_1)^2 + \sum (y_{2j}-\bar{y}_2)^2 + \sum (y_{3j}-\bar{y}_3)^2\\
&= \sum y_{ij}^2 - (4\bar{y}_1^2 + 3\bar{y}_2^2 + 3\bar{y}_3^2)\\
&= \sum y_{ij}^2 - \left[\frac{(\sum y_{1j})^2}{4} + \frac{(\sum y_{2j})^2}{3} + \frac{(\sum y_{3j})^2}{3}\right].
\end{aligned}$$

Interpret each expression in terms of the geometry. The length $s(\mathbf{y})$ gives the scale of $\mathbf{y}$ relative to the subspace $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$.

(iii) The position and reference-point decomposition of Y is

$$Y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \bar{y}_1 & \bar{y}_2 & \bar{y}_3 & s(\mathbf{y}) \end{bmatrix} \begin{bmatrix} \mathbf{v}_1' \\ \mathbf{v}_2' \\ \mathbf{v}_3' \\ \mathbf{d}'(\mathbf{y}) \end{bmatrix}.$$

Check that the reduced structural equation has components

$$\begin{aligned} \bar{y}_1 &= \mu_1 + \sigma\bar{e}_1, \\ \bar{y}_2 &= \mu_2 + \sigma\bar{e}_2, \\ \bar{y}_3 &= \mu_3 + \sigma\bar{e}_3, \\ s(\mathbf{y}) &= \qquad \sigma s(e). \end{aligned}$$

(iv) Show that the positive-lower-triangular and orthogonal factorization of Y is

$$\begin{bmatrix} \sqrt{4} & 0 & 0 & 0 \\ 0 & \sqrt{3} & 0 & 0 \\ 0 & 0 & \sqrt{3} & 0 \\ \sqrt{4}\,\bar{y}_1 & \sqrt{3}\,\bar{y}_2 & \sqrt{3}\,\bar{y}_3 & s(\mathbf{y}) \end{bmatrix}$$

$$\left[\begin{array}{cccc:ccc:ccc} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{y_{11}-\bar{y}_1}{s(\mathbf{y})} & \frac{y_{12}-\bar{y}_1}{s(\mathbf{y})} & \frac{y_{13}-\bar{y}_1}{s(\mathbf{y})} & \frac{y_{14}-\bar{y}_1}{s(\mathbf{y})} & \frac{y_{21}-\bar{y}_2}{s(\mathbf{y})} & \frac{y_{22}-\bar{y}_2}{s(\mathbf{y})} & \frac{y_{23}-\bar{y}_2}{s(\mathbf{y})} & \frac{y_{31}-\bar{y}_3}{s(\mathbf{y})} & \frac{y_{32}-\bar{y}_3}{s(\mathbf{y})} & \frac{y_{33}-\bar{y}_3}{s(\mathbf{y})} \end{array}\right]$$

and check that the analysis-of-variance table is

Source	Dimension	Component	Structure of Component
First	1	$\bar{y}_1^2 4 = \frac{(\sum y_{1j})^2}{4}$	$(\mu_1\sqrt{4} + \sigma\bar{e}_1\sqrt{4})^2$
Second	1	$\bar{y}_2^2 3 = \frac{(\sum y_{2j})^2}{3}$	$(\mu_2\sqrt{3} + \sigma\bar{e}_2\sqrt{3})^2$
Third	1	$\bar{y}_3^2 3 = \frac{(\sum y_{3j})^2}{3}$	$(\mu_3\sqrt{3} + \sigma\bar{e}_3\sqrt{3})^2$
Residual	7	$s^2(\mathbf{y})$	$(\sigma s(\mathbf{e}))^2$
	10	$\sum y_{ij}^2$	

2. (*Continuation*). (i) Show that the invariant differential for the transformation $\tilde{\mathbf{y}} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + m_3\mathbf{v}_3 + c\mathbf{y}$ is

$$\frac{d\mathbf{y}}{s^{10}(\mathbf{y})}$$

as based on Euclidean volume at $\mathbf{d}(\mathbf{y})$. For the left and right transformations described by

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \tilde{m}_1 & \tilde{m}_2 & \tilde{m}_3 & \tilde{c} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ m_1 & m_2 & m_3 & c \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ m_1^* & m_2^* & m_3^* & c^* \end{bmatrix}.$$

show that the left and right invariant differentials on the group are

$$\frac{dm_1\,dm_2\,dm_3\,dc}{c^4}, \qquad \frac{dm_1\,dm_2\,dm_3\,dc}{c}.$$

(ii) Show that the error probability distribution for the reduced model is

$$k(\mathbf{d}) \prod_1^4 f(\bar{e}_1 + sd_{1j}) \prod_1^3 f(\bar{e}_2 + sd_{2j}) \prod_1^3 f(\bar{e}_3 + sd_{3j}) s^6 \, d\bar{e}_1\, d\bar{e}_2\, d\bar{e}_3\, ds,$$

that the marginal error distribution for $t_1 = \bar{e}_1/s$, $t_2 = \bar{e}_2/s$, $t_3 = \bar{e}_3/s$ is

$$k(\mathbf{d}) \int_0^\infty \prod_1^4 f(s(t_1 + d_{1j})) \prod_1^3 f(s(t_2 + d_{2j})) \prod_1^3 f(s(t_3 + d_{3j})) s^9 \, ds \cdot dt_1\, dt_2\, dt_3,$$

and that the marginal error distribution for s is

$$k(\mathbf{d}) \iiint_{-\infty}^{\infty} \prod_1^4 f(\bar{e}_1 + sd_{1j}) \prod_1^3 f(\bar{e}_2 + sd_{2j}) \prod_1^3 f(\bar{e}_3 + sd_{3j}) \, d\bar{e}_1\, d\bar{e}_2\, d\bar{e}_3 \cdot s^6\, ds.$$

(iii) Obtain expressions for the structural distribution for $(\mu_1, \mu_2, \mu_3, \sigma)$, the marginal structural distribution for (μ_1, μ_2, μ_3), and the marginal structural distribution for σ.

(iv) For the case of normal error

$$\prod_1^{10} f(e_{ij}) = (2\pi)^{-5} \exp\{-\tfrac{1}{2}\sum e_{ij}^2\}$$

verify from (ii) that the reduced model can be presented as

$$\begin{aligned} &z_1, \quad z_2, \quad z_3, \quad \chi_7. \\ \bar{y}_1\sqrt{4} &= \mu_1\sqrt{4} + \sigma z_1, \\ \bar{y}_2\sqrt{3} &= \mu_2\sqrt{3} + \sigma z_2, \\ y_3\sqrt{3} &= \mu_2\sqrt{3} + \sigma z_3, \\ s(y) &= \qquad \sigma\chi_7, \end{aligned}$$

of freedom. Also verify from (ii) that the error probability distribution of (t_1, t_2, t_3) is

$$\frac{A_7}{A_{10}} \frac{\sqrt{4}\sqrt{3}\sqrt{3}}{(1 + 4t_1^2 + 3t_2^2 + 3t_3^2)^5}\, dt_1\, dt_2\, dt_3$$

and of the t's individually is

$$\frac{A_7}{A_8} \frac{\sqrt{4}\, dt_1}{(1 + 4t_1^2)^4}, \qquad \frac{A_7}{A_8} \frac{\sqrt{3}\, dt_2}{(1 + 3t_2^2)^4}, \qquad \frac{A_7}{A_8} \frac{\sqrt{3}\, dt_3}{(1 + 3t_3^2)^4}.$$

(v) For the case of normal error give equations that describe the structural distribution of $(\mu_1, \mu_2, \mu_3, \sigma)$ in terms of normal and chi variables, the structural distribution of (μ_1, μ_2, μ_3) in terms of simplified t-variables, and the structural distribution of σ in terms of a chi variable.

3. The model in Problems 1 and 2 could apply equally to a process with stable internal error $f(e)\,de$ and with four response observations under a first set of conditions, three response observations under a second set, and three under a third. If the general response level is unaffected by change of conditions, the measurement model in Chapter One is appropriate. If the general response level is viewed as dependent on the conditions, the model in Problem 1 is appropriate. The general response-level vector in the first case lies in $L(\mathbf{1})$ and in the second case, in $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ with $\mathbf{v}$'s as defined in Problem 1. The change from the one-dimensional subspace $L(\mathbf{1})$ to the three-dimensional subspace $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ requires effectively two additional vectors. The two additional vectors could be chosen from $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, or they could be constructed directly with a view toward orthogonality and ease of interpretation.

$$\prod f(e_{ij}) \prod de_{ij}$$

$$\left[\begin{array}{cccc:ccc:ccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{7} & -\frac{3}{7} & -\frac{3}{7} & -\frac{3}{7} & \frac{4}{7} & \frac{4}{7} & \frac{4}{7} & 0 & 0 & 0 \\ -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & \frac{7}{10} & \frac{7}{10} & \frac{7}{10} \\ y_{11} & y_{12} & y_{13} & y_{14} & y_{21} & y_{22} & y_{23} & y_{31} & y_{32} & y_{33} \end{array}\right]$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \alpha_1 & \alpha_2 & \alpha_3 & \sigma \end{bmatrix} \left[\begin{array}{cccc:ccc:ccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -\frac{3}{7} & -\frac{3}{7} & -\frac{3}{7} & -\frac{3}{7} & \frac{4}{7} & \frac{4}{7} & \frac{4}{7} & 0 & 0 & 0 \\ -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & -\frac{3}{10} & \frac{7}{10} & \frac{7}{10} & \frac{7}{10} \\ e_{11} & e_{12} & e_{13} & e_{14} & e_{21} & e_{22} & e_{23} & e_{31} & e_{32} & e_{33} \end{array}\right].$$

Note that the three structural vectors, to be designated $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$, are mutually orthogonal. The usual conventional has w-vectors obtained by successive orthogonalization of v-vectors. This convention *does not* apply here; the w-vectors are not derived† from the use of two v-vectors in addition to the 1-vector. The w-notation is used merely to suggest a *constructed* orthogonal set.

(i) Verify the interpretations: α_1 is the average general-response level for the 10 performances, α_2 is the general level for the second conditions as it exceeds the general level for the first conditions, α_3 is the general level for the third conditions as it exceeds the average general level for the seven performances with the first two conditions.

† For comparison, construct an orthogonal set by successively orthogonalizing $\mathbf{1}, \mathbf{v}_2, \mathbf{v}_3$.

(ii) Show that the projection of $\mathbf{y}$ into the subspace $L(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$ is

$$a_1(\mathbf{y})\mathbf{w}_1 + a_2(\mathbf{y})\mathbf{w}_2 + a_3(\mathbf{y})\mathbf{w}_3 = (\bar{y}_1, \bar{y}_1, \bar{y}_1, \bar{y}_1, \bar{y}_2, \bar{y}_2, \bar{y}_2, \bar{y}_3, \bar{y}_3, \bar{y}_3),$$

where

$$a_1(\mathbf{y}) = \frac{(\mathbf{w}_1, \mathbf{y})}{(\mathbf{w}_1, \mathbf{w}_1)} \qquad a_2(\mathbf{y}) = \frac{(\mathbf{w}_2, \mathbf{y})}{(\mathbf{w}_2, \mathbf{w}_2)} \qquad a_3(\mathbf{y}) = \frac{(\mathbf{w}_3, \mathbf{y})}{(\mathbf{w}_3, \mathbf{w}_3)}$$

$$= \bar{y} = \frac{\sum y_{ij}}{10}, \qquad = \bar{y}_2 - \bar{y}_1, \qquad = \bar{y}_3 - \frac{4\bar{y}_1 + 3\bar{y}_2}{7}.$$

Check that the squared-residual-length is

$$\begin{aligned} s^2(\mathbf{y}) &= \sum (y_{1j} - \bar{y}_1)^2 + \sum (y_{2j} - \bar{y}_2)^2 + \sum (y_{3j} - \bar{y}_3)^2 \\ &= \sum y_{ij}^2 - 10\bar{y}^2 - \frac{12}{7}(\bar{y}_2 - \bar{y}_1)^2 - \frac{21}{10}\left(\bar{y}_3 - \frac{4\bar{y}_1 + 3\bar{y}_3}{7}\right)^2. \end{aligned}$$

The reduced structural equation has components

$$\begin{aligned} a_1(\mathbf{y}) &= \alpha_1 + \sigma a_1(\mathbf{e}), \\ a_2(\mathbf{y}) &= \alpha_2 + \sigma a_2(\mathbf{e}), \\ a_3(\mathbf{y}) &= \alpha_3 + \sigma a_3(\mathbf{e}), \\ s(\mathbf{y}) &= \quad \sigma s(\mathbf{e}). \end{aligned}$$

(iii) Show that the positive-lower-triangular and orthogonal factorization of Y is

$$\begin{bmatrix} \sqrt{10} & 0 & 0 & 0 \\ 0 & \sqrt{\frac{12}{7}} & 0 & 0 \\ 0 & 0 & \sqrt{\frac{21}{10}} & 0 \\ \sqrt{10}\,\bar{y} & \sqrt{\frac{12}{7}}\, a_2(\mathbf{y}) & \sqrt{\frac{21}{10}}\, a_3(\mathbf{y}) & s(\mathbf{y}) \end{bmatrix}$$

$$\cdot \left[\begin{array}{cccc:ccc:ccc} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{-3}{\sqrt{84}} & \frac{-3}{\sqrt{84}} & \frac{-3}{\sqrt{84}} & \frac{-3}{\sqrt{84}} & \frac{4}{\sqrt{84}} & \frac{4}{\sqrt{84}} & \frac{4}{\sqrt{84}} & 0 & 0 & 0 \\ \frac{-3}{\sqrt{210}} & \frac{-3}{\sqrt{210}} & \frac{-3}{\sqrt{210}} & \frac{-3}{\sqrt{210}} & \frac{-3}{\sqrt{210}} & \frac{-3}{\sqrt{210}} & \frac{-3}{\sqrt{210}} & \frac{7}{\sqrt{210}} & \frac{7}{\sqrt{210}} & \frac{7}{\sqrt{210}} \\ d_{11} & d_{12} & d_{13} & d_{14} & d_{21} & d_{22} & d_{23} & d_{31} & d_{32} & d_{33} \end{array}\right].$$

Check that the analysis-of-variance table is

Source	Dimension	Component	Structure of Component
$\mathbf{w}_1$	1	$\bar{y}^2\, 10$	$(\alpha_1\sqrt{10} + \sigma\bar{e}\sqrt{10})^2$
$\mathbf{w}_2$	1	$(\bar{y}_2 - \bar{y}_1)^2\, \frac{12}{7}$	$(\alpha_2\sqrt{\frac{12}{7}} + \sigma(\bar{e}_2 - \bar{e}_1)\sqrt{\frac{12}{7}})^2$
$\mathbf{w}_3$	1	$\left(\bar{y}_3 - \frac{4\bar{y}_1 + 3\bar{y}_2}{7}\right)^2 \frac{21}{10}$	$\left(\alpha_3\sqrt{\frac{21}{10}} + \sigma\left(\bar{e}_3 - \frac{4\bar{e}_1 + 3\bar{e}_2}{7}\right)\sqrt{\frac{21}{10}}\right)^2$
Residual	$\frac{7}{10}$	$\frac{s^2(\mathbf{y}) = \sum (y_{ij} - \bar{y}_i)^2}{\sum y_{ij}^2}$	$(\sigma s(\mathbf{e}))^2$

Note that the first three components in Problem 1 have, in effect, been combined and new first components formed: one set of orthogonal axes in $L(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = L(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$ has been replaced by another set of orthogonal axes.

4. (*Continuation*). (i) Check that the usual error probability distributions for the reduced model are

$$k(\mathbf{d}) \prod f(a_1 w_{1ij} + a_2 w_{2ij} + a_3 w_{3ij} + s d_{ij}) s^6 da\, ds,$$

$$k(\mathbf{d}) \int_0^\infty \prod f(s(\bar{t}_1 w_{1ij} + \bar{t}_2 w_{2ij} + \bar{t}_3 w_{3ij} + d_{ij})) s^9\, ds \cdot d\bar{t},$$

$$k(\mathbf{d}) \iiint_{-\infty}^{\infty} \prod f(a_1 w_{1ij} + a_2 w_{2ij} + a_3 w_{3ij} + s d_{ij}) da \cdot s^6\, ds.$$

(ii) For the case of normal error

$$\prod_1^{10} f(e_{ij}) = (2\pi)^{-5} \exp\{-\tfrac{1}{2}\textstyle\sum e_{ij}^2\}$$

verify that the reduced model can be represented as

$$z_1, \quad z_2, \quad z_3, \quad \chi_7,$$

$$a_1(\mathbf{y})\sqrt{10} = \alpha_1\sqrt{10} + \sigma z_1,$$

$$a_2(\mathbf{y})\sqrt{\tfrac{12}{7}} = \alpha_2\sqrt{\tfrac{12}{7}} + \sigma z_2,$$

$$a_3(\mathbf{y})\sqrt{\tfrac{21}{10}} = \alpha_3\sqrt{\tfrac{21}{10}} + \sigma z_3,$$

$$s(\mathbf{y}) = \sigma\chi_7.$$

Also verify that the error probability distribution of (t_1, t_2, t_3) is

$$\frac{A_7}{A_{10}} \frac{\sqrt{10}\,\sqrt{12/7}\,\sqrt{21/10}}{(1 + 10\bar{t}_1^2 + (12/7)\bar{t}_2^2 + (21/10)\bar{t}_3^2)^5} d\bar{t}_1\, d\bar{t}_2\, d\bar{t}_3,$$

and of $\bar{t}_3$, for example, is

$$\frac{A_7}{A_8} \frac{\sqrt{21/10}}{(1 + (21/10)\bar{t}_3^2)^4} d\bar{t}_3.$$

(iii) For the case of normal error give equations that describe the structural distributions of $(\alpha_1, \alpha_2, \alpha_3, \sigma)$, $(\alpha_1, \alpha_2, \alpha_3)$, and σ (use normal, chi, and simplified t-variables).

5. Consider a process with stable error $f(e)de$ and a response y whose general level is known to depend linearly on a controllable variable x. For five response observations the model would be

$$\prod_1^5 f(e_i) \prod de_i,$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 & x_5 \\ y_1 & y_2 & y_3 & y_4 & y_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_0 & \beta_1 & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 & x_5 \\ e_1 & e_2 & e_3 & e_4 & e_5 \end{bmatrix}.$$

(i) Show that the projection of $\mathbf{y}$ into the subspace $L(\mathbf{1}, \mathbf{x})$ is $b_0(\mathbf{y})\mathbf{1} + b_1(\mathbf{y})\mathbf{x}$, where

$$b_1(\mathbf{y}) = \frac{\sum (x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

$$b_0(\mathbf{y}) = \bar{y} - b_1(\mathbf{y})\bar{x};$$

and show that the squared residual length is

$$\begin{aligned} s^2(\mathbf{y}) &= \sum (y_i - b_0(\mathbf{y}) - b_1(\mathbf{y})x_i)^2 \\ &= \sum (y_i - \bar{y} - b_1(\mathbf{y})(x_i - \bar{x}))^2 \\ &= \sum y_i^2 - \frac{(\sum y_i)^2}{n} - \frac{(\sum (x_i - \bar{x})y_i)^2}{\sum (x_i - \bar{x})^2}. \end{aligned}$$

(ii) The reduced structural equation has components

$$b_0(\mathbf{y}) = \beta_0 + \sigma b_0(\mathbf{e})$$

$$b_1(\mathbf{y}) + \beta_1 = \sigma b_1(\mathbf{e})$$

$$s(\mathbf{y}) = \sigma s(\mathbf{e}).$$

Give expressions for the error probability distribution of $(b_0(\mathbf{e}), b_1(\mathbf{e}), s(\mathbf{e}))$, $(t_0(\mathbf{e}), t_1(\mathbf{e}))$, and $s(\mathbf{e})$.

(iii) For the case of normal error

$$\prod_1^5 f(e_i) = (2\pi)^{-5/2} \exp\{-\textstyle\sum e_i^2\}$$

record the error probability elements for $(b_0(\mathbf{e}), b_1(\mathbf{e}), s(\mathbf{e}))$, $(t_0(\mathbf{e}), t_1(\mathbf{e}))$, and $s(\mathbf{e})$. (The quadratic expressions have cross-product terms: the general-case nonorthogonality of $\mathbf{1}$ and $\mathbf{x}$.)

6. (*Continuation*). Consider the preceding model in orthogonal form

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ x_1 - \bar{x} & x_2 - \bar{x} & x_3 - \bar{x} & x_4 - \bar{x} & x_5 - \bar{x} \\ y_1 & y_2 & y_3 & y_4 & y_5 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_0 & \alpha_1 & \sigma \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ x_1 - \bar{x} & x_2 - \bar{x} & x_3 - \bar{x} & x_4 - \bar{x} & x_5 - \bar{x} \\ e_1 & e_2 & e_3 & e_4 & e_5 \end{bmatrix}.$$

(i) Express the α's in terms of the β's with P^{-1} and the β's in terms of the α's with P.

(ii) Show that the projection of $\mathbf{y}$ into the subspace $L(\mathbf{w}_1, \mathbf{w}_2)$ is

$$a_0(\mathbf{y})\mathbf{1} + a_1(\mathbf{y})(\mathbf{x} - \bar{x}\mathbf{1}),$$

where

$$a_0(\mathbf{y}) = \bar{y}$$

$$a_1(\mathbf{y}) = \frac{\sum (x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

and show that the squared residual length is

$$
\begin{aligned}
s^2(\mathrm{y}) &= \sum (y_i - \bar{y} - a_1(\mathrm{y})(x_i - \bar{x}))^2 \\
&= \sum y_i^2 - \frac{(\sum y_i)^2}{n} - \frac{(\sum (x_i - \bar{x})y_i)^2}{\sum (x_i - \bar{x})^2}.
\end{aligned}
$$

Note that the highest order coefficients in the two forms for the model are equal: $\beta_1 = \alpha_1$; $b_1(\mathrm{y}) = a_1(\mathrm{y})$. The reduced structural equation has components

$$
\begin{aligned}
a_0(\mathrm{y}) &= \alpha_0 + \sigma a_0(\mathrm{e}), \\
a_1(\mathrm{y}) &= \alpha_1 + \sigma a_1(\mathrm{e}), \\
s(\mathrm{y}) &= \sigma s(\mathrm{e}).
\end{aligned}
$$

(iii) Show that the positive-lower-triangular and orthogonal factorization of $\bar{Y}$ is

$$
\begin{bmatrix}
\sqrt{5} & 0 & 0 \\
0 & \sqrt{\sum (x_i - \bar{x})^2} & 0 \\
\sqrt{5}\,\bar{y} & \sqrt{\sum (x_i - \bar{x})^2}\,a_1(\mathrm{y}) & s(\mathrm{y})
\end{bmatrix}
\cdot
$$

$$
\begin{bmatrix}
\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\
\frac{x_1 - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}} & \frac{x_2 - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}} & \frac{x_3 - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}} & \frac{x_4 - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}} & \frac{x_5 - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}} \\
d_1(\mathrm{y}) & d_2(\mathrm{y}) & d_3(\mathrm{y}) & d_4(\mathrm{y}) & d_5(\mathrm{y})
\end{bmatrix}.
$$

Check that the analysis-of-variance table is

Source	Dimension	Component	Structure of Component
Constant	1	$\bar{y}^2 5$	$(\alpha_0\sqrt{5} + \sigma\bar{e}\sqrt{5})^2$
Linear	1	$a_1^2(\mathrm{y})\sum (x_i - \bar{x})^2$	$\left(\alpha_1\sqrt{\sum (x_i - \bar{x})^2} + \sigma a_1(\mathrm{e})\sqrt{\sum (x_i - \bar{x})^2}\right)^2$
Residual	3	$s^2(\mathrm{y})$	$(\sigma\chi_3)^2$
	5	$\sum y_i^2$	

(iv) Give expressions for the error probability distribution of $(a_0(e), a_1(e), s(e))$, $(\bar{t}_0(e), \bar{t}_1(e))$, and $s(e)$.

(v) For the case of normal error verify that the reduced model can be presented as

$$
z_0, \quad z_1, \quad \chi_3,
$$

$$
\begin{aligned}
\bar{y}\sqrt{5} &= \alpha_0\sqrt{5} &&+ \sigma z_0, \\
a_1(\mathrm{y})\sqrt{\sum (x_i - \bar{x})^2} &= \alpha_1\sqrt{\sum (x_i - \bar{x})^2} &&+ \sigma z_1, \\
s(\mathrm{y}) &= &&\sigma\chi_3.
\end{aligned}
$$

Also verify that the error probability distribution of $(\bar{t}_0(\mathrm{e}), \bar{t}_1(\mathrm{e}))$ is

$$
\frac{A_3}{A_5}\,\frac{\sqrt{5}\sqrt{\sum (x_i - \bar{x})^2}}{(1 + 5\bar{t}_0^2 + \sum (x_i - \bar{x})^2\bar{t}_1^2)^{5/2}}\,d\bar{t}_0\,d\bar{t}_1
$$

and of $\bar{t}_1(\mathrm{e})$, for example, is

$$
\frac{A_3}{A_4}\,\frac{\sqrt{\sum (x_i - \bar{x})^2}}{(1 + \sum (x_i - \bar{x})^2\bar{t}_1^2)^2}\,d\bar{t}_1.
$$

(vi) For the case of normal error give equations that describe the structural distributions of $(\alpha_0, \alpha_1, \sigma)$, (α_0, α_1), α_1, and σ (use normal, chi, and simplified t-variables).

Note: On the assumption that there is *no* dependence on the controllable variable, the measurement model of Chapter One is appropriate. On the assumption that the dependence is *linear*, the type of model in Problems 5 and 6 is appropriate. Now suppose that several observations are taken at each level of the controllable variable. On the assumption that the dependence is of *unknown form*, the type of model in Problems 1, 2, 3, and 4 is appropriate. An analysis-of-variance table for such a succession of models can be calculated by using the results from the various problems; tests and structural distributions can be obtained from a particular model. For examples, see Problems 13, 20, 21, and 23.

7. Consider a process with a known error distribution and suppose that 12 observations are taken, two at each combination of levels A_1, A_2 for a first factor A affecting the process and levels B_1, B_2, B_3 of a second factor B affecting the process:

	B_1		B_2		B_3	
A_1	y_{111}	y_{112}	y_{121}	y_{122}	y_{131}	y_{132}
A_2	y_{211}	y_{212}	y_{221}	y_{222}	y_{231}	y_{232}

If the factors were known not to affect the general response level, the measurement model in Chapter One would be appropriate.

If the factor B were known *not* to affect the general response level, the kind of model in Problems 1, 2, 3, 4 would be applicable (two levels for the factor A and, accordingly, two structural vectors):

$$
\prod f(e_{ijs}) \prod de_{ijs},
$$

$$
\left[\begin{array}{cc:cc:cc:cc:cc:cc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\
y_{111} & y_{112} & y_{121} & y_{122} & y_{131} & y_{132} & y_{211} & y_{212} & y_{221} & y_{222} & y_{231} & y_{232}
\end{array}\right]
$$

$$
= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \rho & \sigma \end{bmatrix}
\begin{bmatrix} 1 & \cdots & 1 \\ -1 & \cdots & 1 \\ e_{111} & \cdots & e_{232} \end{bmatrix}.
$$

Similarly, if the factor A were known not to affect the general response level, the kind of model in Problems 1, 2, 3, and 4 would be applicable (three levels for the factor B and,

accordingly, three structural vectors):

$$\prod f(e_{ijs}) \prod de_{ijs},$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\ -1 & -1 & -1 & -1 & 2 & 2 & -1 & -1 & -1 & -1 & 2 & 2 \\ y_{111} & y_{112} & y_{121} & y_{122} & y_{131} & y_{132} & y_{211} & y_{212} & y_{221} & y_{222} & y_{231} & y_{232} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & 0 \\ 0 & 1 & & \\ 0 & 0 & 1 & \\ \mu & \gamma_1 & \gamma_2 & \sigma \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \\ -1 \cdots 0 \\ -1 \cdots 2 \\ e_{111} \cdots e_{232} \end{bmatrix}.$$

Note that the lengths of the structural vectors have been chosen to avoid fractions.

More generally, perhaps the factors are known to have possible effects on the general response level but only in an *additive* manner: a change in factor A changes the general level by the same amount at each level of B and a change in factor B changes the general level by the same amount at each level of A. A combination of the preceding two models is then appropriate.

$$\prod f(e_{ijs}) \prod de_{ijs}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\ -1 & -1 & -1 & -1 & 2 & 2 & -1 & -1 & -1 & -1 & 2 & 2 \\ y_{111} & y_{112} & y_{121} & y_{122} & y_{131} & y_{132} & y_{211} & y_{212} & y_{221} & y_{222} & y_{231} & y_{232} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & & 0 \\ 0 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & 1 & \\ \mu & \rho & \gamma_1 & \gamma_2 & \sigma \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \\ -1 \cdots 1 \\ -1 \cdots 0 \\ -1 \cdots 2 \\ e_{111} \cdots e_{232} \end{bmatrix}.$$

Note that the four structural vectors are mutually orthogonal: in particular, the second vector describing row differences is orthogonal to the third and fourth vectors describing column differences.

And more generally still, if the factors are known to have possible effects on the general response level unrestricted by additivity, the kind of model in Problems 1 and 2 would be applicable with six general levels given by the table

	B_1	B_2	B_3
A_1	μ_{11}	μ_{12}	μ_{13}
A_2	μ_{21}	μ_{22}	μ_{23}

and with a corresponding six orthogonal structural vectors. Alternatively, the model can be structured by adding two vectors to the four structural vectors in the additive model:

$$\prod f(e_{ijs}) \prod de_{ijs},$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\ -1 & -1 & -1 & -1 & 2 & 2 & -1 & -1 & -1 & -1 & 2 & 2 \\ 1 & 1 & -1 & -1 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & -2 & -2 & -1 & -1 & -1 & -1 & 2 & 2 \\ y_{111} & y_{112} & y_{121} & y_{122} & y_{131} & y_{132} & y_{211} & y_{212} & y_{221} & y_{222} & y_{231} & y_{232} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & & & & 0 \\ 0 & 1 & & & & & \\ 0 & 0 & 1 & & & & \\ 0 & 0 & 0 & 1 & & & \\ 0 & 0 & 0 & 0 & 1 & & \\ 0 & 0 & 0 & 0 & 0 & 1 & \\ \mu & \rho & \gamma_1 & \gamma_2 & \alpha_1 & \alpha_2 & \sigma \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \\ -1 \cdots 1 \\ -1 \cdots 0 \\ -1 \cdots 2 \\ 1 \cdots 0 \\ 1 \cdots 2 \\ e_{111} \cdots e_{232} \end{bmatrix}.$$

(i) Check the mutual orthogonality of the structural vectors in the final model. The fifth vector is obtained by multiplying corresponding elements in the second and third vectors, the sixth vector, by multiplying corresponding elements in the second and fourth vectors. Justify this procedure in terms of the interpretation of vectors as describing row differences and column differences.

(ii) The projection of $\mathbf{y}$ into the six dimensional subspace determined by the structural vectors $\mathbf{w}_1, \ldots, \mathbf{w}_6$ in the final model can be written

$$m\mathbf{w}_1 + r\mathbf{w}_2 + c_1\mathbf{w}_3 + c_2\mathbf{w}_4 + a_1\mathbf{w}_5 + a_2\mathbf{w}_6.$$

Give expressions for the coefficients in terms of averages such as

$$\bar{y}_{11\cdot} = \frac{\sum_s y_{11s}}{2}, \qquad \bar{y}_{\cdot 2\cdot} = \frac{\sum_{is} y_{i2s}}{4}.$$

(iii) Determine the positive-lower-triangular and orthogonal factorization of the matrix Y; use the notation $m(\mathbf{y}), \ldots, a_2(\mathbf{y})$. Record the analysis-of-variance table.

8. (*Continuation*). For the case of normal error

$$\prod f(e_{ijs}) = (2\pi)^{-6} \exp\{-\tfrac{1}{2}\textstyle\sum e_{ijs}^2\}$$

express the reduced model in terms of normal and chi variables (cf. Problems 4 and 6).

9. Eleven pieces of material were sampled at random from a lot; five, chosen at random, were subjected to a first treatment, and the remaining six to a second treatment. Suppose the

regression model with normal error is applicable. Let μ_1 be the level for the first treatment and μ_2 for the second treatment; and suppose the observations yield

$$\bar{y}_1 = 0.275, \qquad \bar{y}_2 = 0.293,$$
$$s_{y_1}^2 = 0.00045, \qquad s_{y_2}^2 = 0.00039.$$

Derive the structural distribution for $\mu_2 - \mu_1$ (note that *variances* are recorded).

10. Show that the location group L is a normal subgroup of the regression-scale group G (Section 7).

11. Measurements are made to estimate a period of oscillation. Let y_0 be the measured time when the oscillation is in a certain phase, and $y_1, \ldots, y_n$ be the measured times for successive recurrence to that phase. Suppose the measurement error is normal: then the regression model with equation

$$y_i = \beta_1 + \beta_2 i + \sigma e_i$$

is applicable; β_2 is the period of oscillation. Show that the structural distribution for β_2 is located at

$$\beta_2 = \frac{\sum y_i(i - n/2)}{\sum (i - n/2)^2}$$

and has t-form on $n - 1$ degrees of freedom; obtain an expression for the scaling of the t-distribution.

12. Consider the regression model with normal error, with structural relationship

$$y = \beta_1 1 + \beta_2 x + \beta_3 x^2 + \sigma e,$$

and observations

x	2.6	2.7	2.8	2.9	3.1
y	12.1	12.5	12.7	13.0	13.5.

on the response y corresponding to the controllable variable x.

(i) Test the *hypothesis* $\beta_3 = 0$.
(ii) On the assumption that $\beta_3 = 0$ derive the structural distribution for β_2.

13 (*Continuation*). Two determinations of the response y were made at each of five levels of a controllable variable x:

x	2.6	2.7	2.8	2.9	3.1
y	12.2	12.5	12.5	13.1	13.5
	12.0	12.5	12.9	12.9	13.5.

Suppose the regression model is applicable: normal error and response levels μ_1, μ_2, μ_3, μ_4, μ_5 for the five levels of controllable variable. Calculate the analysis-of-variance table with entries for: mean; linear dependence on x; quadratic dependence on x; other dependence on x; residual. Test the hypothesis that the dependence is at most quadratic.

14. Use the orthogonal basis to show that

$$|\mathbf{y} - \sum_1^r b_u^{(r)}(\mathbf{y})\mathbf{v}_u|^2 = |\mathbf{y} - \sum_1^{r+1} b_u^{(r+1)}(\mathbf{y})\mathbf{v}_u|^2 + (b_{r+1}^{(r+1)}(\mathbf{y})\,|\mathbf{w}_{r+1}|)^2.$$

15. Use the orthogonal basis to show that

$$\begin{aligned}(\mathbf{y}_1 - \sum_1^r b_u^{(r)}(\mathbf{y}_1)\mathbf{v}_u, \mathbf{y}_2 - \sum_1^r b_u^{(r)}(\mathbf{y}_2)\mathbf{v}_u) &= (\mathbf{y}_1 - \sum_1^{r+1} b_u^{(r+1)}(\mathbf{y}_1)\mathbf{v}_u, \mathbf{y}_2 - \sum_1^{r+1} b_u^{(r+1)}(\mathbf{y}_2)\mathbf{v}_u) \\ &\quad + b_{r+1}^{(r+1)}(\mathbf{y}_1)b_{r+1}^{(r+1)}(\mathbf{y}_2)(\mathbf{w}_{r+1}, \mathbf{w}_{r+1}).\end{aligned}$$

16. Show that

$$\begin{aligned}(\mathbf{y}_1 - \sum_1^r b_u(\mathbf{y}_1)\mathbf{v}_u, \mathbf{y}_2 - \sum_1^r b_u(\mathbf{y}_2)\mathbf{v}_u) &= (\mathbf{y}_1, \mathbf{y}_2) - \sum_u b_u(\mathbf{y}_1)(\mathbf{v}_u, \mathbf{y}_2) \\ &= (\mathbf{y}_1, \mathbf{y}_2) - \sum_u b_u(\mathbf{y}_2)(\mathbf{v}_u, \mathbf{y}_1),\end{aligned}$$

$$|\mathbf{y} - \sum b_u(\mathbf{y})\mathbf{v}_u|^2 = |\mathbf{y}|^2 - \sum_u b_u(\mathbf{y})(\mathbf{v}_u, \mathbf{y}).$$

17. Consider the regression model but with known error scaling:

$$f(E)\,dE = \prod_1^n f(e_i) \prod_1^n de_i,$$
$$Y = \theta E,$$

where

$$Y = \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{y}' \end{bmatrix}, \qquad E = \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{e}' \end{bmatrix}, \qquad \theta = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ \beta_1 & & \cdots & \beta_r & 1 \end{bmatrix},$$

the simple regression model. Assume that $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent and θ is an element of the regression group

$$G = \left\{ \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ 0 & & & 1 & 0 \\ b_1 & & \cdots & b_r & 1 \end{bmatrix} : \; -\infty < b_u < \infty \right\}.$$

(i) Determine a transformation variable and a reference point (Notation of Section 3; pattern of Section 3 in Chapter One). Check that the model is a structural model.
(ii) Derive the invariant differentials and the modular function.
(iii) Derive the distribution for error position and the structural distribution for θ.

18 (*Continuation*). Consider the simple regression model with normal component error:

$$f(e)\,de = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left\{-\frac{1}{2}\frac{e^2}{\sigma_0^2}\right\} de.$$

Derive the distribution of error position: (a) using an orthogonal basis $\mathbf{w}_1, \ldots, \mathbf{w}_r$, and (b) using the given basis $\mathbf{v}_1, \ldots, \mathbf{v}_r$, where

$$W = PV, \qquad V = P^{-1}W.$$

19. Suppose levels $A_1, \ldots, A_a$ are chosen for a factor A. Let y_{is} be the sth response at level A_i:

$$\mathbf{y} = \begin{bmatrix} y_{11} & \cdots & y_{1n_1} \\ y_{21} & \cdots & y_{2n_2} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ y_{a1} & \cdots & y_{an_a} \end{bmatrix} \begin{array}{l} (1) \\ (2) \\ \\ \\ \\ (a) \end{array} = (y_{is});$$

the *vector* is recorded in a convenient *array* with different rows containing responses for different levels; $N = \sum n_i$; $\mathbf{y} \in R^N$. Let $\mathbf{v}_0$ be the $\mathbf{1}$-vector:

$$\mathbf{v}_0 = \begin{bmatrix} 1 & \cdots & 1 \\ 1 & \cdots 1 & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ 1 & \cdots & 1 \end{bmatrix};$$

and let $\mathbf{v}_i$ be an indicator vector for the level A_i:

$$\mathbf{v}_i = \begin{bmatrix} 0 & \cdots & 0 \\ \cdot & & \\ \cdot & & \\ 0 & & \\ 1 & \cdots 1 & \\ 0 & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ 0 & \cdots & 0 \end{bmatrix} \begin{array}{l} (1) \\ \\ \\ (i-1) \\ (i) \\ (i+1) \\ \\ \\ \\ (a) \end{array}$$

(i) In the table check the following entries: ordering of subspaces by "contained in" ($L(\mathbf{v}_0) \subset L(\mathbf{v}_1, \ldots, \mathbf{v}_a) \subset R^N$); dimension of subspace; projection into subspace; squared length of projection.

	Space	Dimension	Projection	Squared Length
Mean	$L(\mathbf{v}_0)$	1	$(\bar{y})$	$N\bar{y}^2 = \boxed{(\sum y_{is})^2/N}$
Factor A	$L(\mathbf{v}_1, \ldots, \mathbf{v}_a)$	a	$(\bar{y}_{i\cdot})$	$\sum n_i\bar{y}_{i\cdot}^2 = \boxed{\sum_i \frac{\left(\sum_s y_{is}\right)^2}{n_i}}$
	R^N	N	(y_{is})	$\boxed{\sum y_{is}^2}$

NOTATION. $\bar{y} = \sum y_{is}/N$; $\bar{y}_{i\cdot} = \sum_s y_{is}/n_i$.

(ii) Justify the elements in the corresponding analysis-of-variance table (regression model with normal error).

Source	Dimension	Component	Structure of Component
Mean	1	$N\bar{y}^2$	$(\sqrt{N}\mu + \sigma z_0)^2$
Factor A	$a - 1$	$\sum n_i\bar{y}_{i\cdot}^2 - N\bar{y}^2 = \sum n_i(\bar{y}_{i\cdot} - \bar{y})^2$	$(\sqrt{\sum n_i(\mu_i - \mu)^2} + \sigma z_1)^2 + \sigma^2\chi_{a-2}^2$
Residual	$N - a$	$\sum y_{is}^2 - \sum n_i\bar{y}_{i\cdot}^2 = \sum (y_{is} - \bar{y}_{i\cdot})^2$	$\sigma^2\chi_{N-a}^2$
Total	N	$\sum y_{is}^2$	

NOTATION. $\mu = \Sigma\, n_i\mu_i/N$; z's are independent standard normal, χ's are independent chi-variables with degrees of freedom as subscribed.

20. Determinations were made on the yield using three methods of catalyzing a chemical process:

I	47.2	49.8	48.5	48.7	
II	50.1	49.3	51.5	50.9	
III	49.1	53.2	51.2	52.8	52.3.

Suppose that the regression model is applicable: normal error and levels μ_1, μ_2, μ_3 for the three methods.

(i) Calculate the analysis-of-variance table by using the enboxed-expressions in Problem 19.

(ii) Derive the structural distribution for $\mu_3 - (\mu_1 + \mu_2)/2$; for $\mu_2 - \mu_1$.

21. Three chemists do three, five, four determinations on the chemical content of a mixture:

I	2.6	2.9	2.8		
II	3.1	3.0	3.3	2.9	3.3
III	2.9	3.2	3.0	3.1	

Suppose that the regression model is applicable: normal error and levels μ_1, μ_2, μ_3 for the three chemists. Calculate the analysis-of-variance table and test the *hypothesis* $\mu_1 = \mu_2 = \mu_3$ (the three chemists are consistent in their measurement levels).

22. Suppose levels $A_1, \ldots, A_a$ are chosen for a factor A, and levels $B_1, \ldots, B_b$ for a factor B. Let y_{ijs} be the sth response at level A_i for A and level B_j for B:

$$\mathbf{y} = \begin{bmatrix} y_{111} \cdots y_{11n} & \cdots & y_{1b1} \cdots y_{1bn} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ y_{a11} \cdots y_{a1n} & \cdots & y_{ab1} \cdots y_{abn} \end{bmatrix} = (y_{ijs});$$

the *vector* is recorded in a convenient *array* separating levels of A by rows and levels of B by columns; $N = abn$. Let $\mathbf{v}_{00}$ be the 1-vector:

$$\mathbf{v}_{00} = \begin{bmatrix} 1 \cdots 1 & \cdots & 1 \cdots 1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 1 \cdots 1 & \cdots & 1 \cdots 1 \end{bmatrix}.$$

Let $\mathbf{v}_{i0}$ be an indicator vector for the level A_i:

$$\mathbf{v}_{i0} = \begin{array}{c} \begin{bmatrix} 0 \cdots 0 & \cdots & 0 \cdots 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 0 & & 0 \\ 1 \cdots 1 & \cdots & 1 \cdots 1 \\ 0 & & 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 0 \cdots 0 & \cdots & 0 \cdots 0 \end{bmatrix} \end{array} \begin{array}{l} (1) \\ \\ \\ \\ \\ (i). \\ \\ \\ \\ \\ (a) \end{array}$$

Let $\mathbf{v}_{0j}$ be an indicator vector for the level B_j:

$$\mathbf{v}_{0j} = \begin{array}{c} \begin{array}{ccc} (1) & (j) & (b) \end{array} \\ \begin{bmatrix} 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ \cdot & & & & & & & & & & \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \end{bmatrix} \end{array}.$$

Let $\mathbf{v}_{ij}$ be an indicator vector for the combination A_iB_j (zeros everywhere except for coordinates corresponding to A_iB_j). This is a *two-factor factorial design*.

(i) Check the ordering ($L' \subset L''$) by "contained-in." Check that the extensions

$$\begin{tikzcd}
& L(\mathbf{v}_{00}) \arrow[dl, dash, "\supset"'] \arrow[dr, dash, "\subset"] \arrow[d, phantom, "\perp"] & \\
L(\mathbf{v}_{10}, \ldots, \mathbf{v}_{a0}) \arrow[dr, dash, "\subset"'] & & L(\mathbf{v}_{01}, \ldots, \mathbf{v}_{0b}) \arrow[dl, dash, "\supset"] \\
& L(\mathbf{v}_{ij}\colon \text{all } i, j) \arrow[d, dash, "\cap"] & \\
& R^N &
\end{tikzcd}$$

from $L(\mathbf{v}_{00})$ are in orthogonal directions ($\perp$ as indicated): a consequence of the same number of observations at each combination A_iB_j of levels.

(ii) In the table check the following entries: dimension of subspace; projection into subspace; squared length of projection.

	Space	Dimension	Projection	Squared Length
Mean	$L(\mathbf{v}_{00})$	1	$(\bar{y}_{\cdots})$	$abn\bar{y}^2_{\cdots} = \boxed{(\sum y_{ijs})^2/N}$
A	$L(\mathbf{v}_{i0}\text{:all } i)$	a	$(\bar{y}_{i\cdot\cdot})$	$bn \sum_i \bar{y}^2_{i\cdot\cdot} = \boxed{\sum_i \left(\sum_{js} y_{ijs}\right)^2 \Big/ bn}$
B	$L(\mathbf{v}_{0j}\text{:all } j)$	b	$(\bar{y}_{\cdot j \cdot})$	$an \sum_j \bar{y}^2_{\cdot j \cdot} = \boxed{\sum_j \left(\sum_{is} y_{ijs}\right)^2 \Big/ an}$
$A \times B$	$L(\mathbf{v}_{ij}\text{:all } i, j)$	ab	$(\bar{y}_{ij\cdot})$	$n \sum_{ij} \bar{y}^2_{ij\cdot} = \boxed{\sum_{ij} \left(\sum_s y_{ijs}\right)^2 \Big/ n}$
	R^N	abn	(y_{ijs})	$\boxed{\sum y^2_{ijs}}$

NOTATION. $\bar{y}_{\cdots} = \sum y_{ijs}/N$; $\bar{y}_{i\cdot\cdot} = \sum_{js} y_{ijs}/bn$; $\bar{y}_{ij\cdot} = \sum_s y_{ijs}/n$.

(iii) Justify the elements in the corresponding analysis-of-variance table (regression model with normal error).

Source	Dimension	Component	Structure of Component
Mean	1	$N\bar{y}^2_{\cdots}$	$(\sqrt{N}\mu + \sigma z_0)^2$
A	$a - 1$	$bn \sum_i \bar{y}^2_{i\cdot\cdot} - N\bar{y}^2$	
		$= bn \sum_i (\bar{y}_{i\cdot\cdot} - \bar{y})^2$	
B	$b - 1$	$an \sum_j \bar{y}^2_{\cdot j\cdot} - N\bar{y}^2$	
		$= an \sum_j (\bar{y}_{\cdot j\cdot} - \bar{y})^2$	
$A \times B$	$(a - 1)(b - 1)$	$n \sum_{ij} \bar{y}^2_{ij\cdot}$ − (preceding entries)	
		$= n \sum_{ij} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y})^2$	
Residual	$ab(n - 1)$	$\sum y^2_{ijs}$ − (preceding entries)	$\sigma^2\chi^2_{ab(n-1)}$
		$= \sum (y_{ijs} - \bar{y}_{ij\cdot})^2$	
	abn	$\sum y^2_{ijs}$	

(iv) Derive expressions for the missing entries under "Structure of Component."

23. A factor A (temperature) is given three levels; a factor B (pressure) is given two levels;

two determinations are made at each combination of levels:

		B: 18	B: 20
A	73	14.2	18.3
		14.3	18.0
	74	17.4	21.4
		17.6	21.0
	75	20.7	23.8
		20.4	23.9

Suppose the regression model is applicable: normal error and level μ_{ij} for the combination A_iB_j.

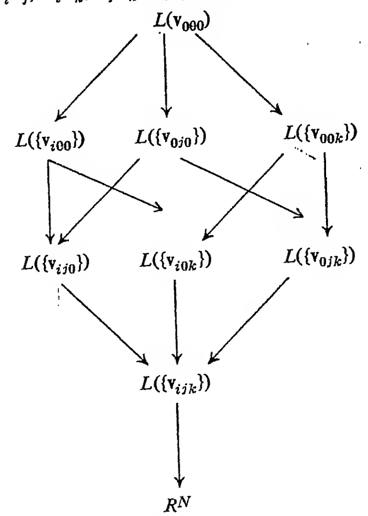
(i) Calculate the analysis-of-variance table by using the enboxed expressions of Problem 22.

(ii) Test the hypothesis that the effects of factors A and B are additive: $\mu_{ij} = \mu + \delta_{i\cdot} + \delta_{\cdot j}$ (i.e., no *interaction* $A \times B$ for factors A and B).

NOTATION.

$$\delta_{i\cdot} = \mu_{i\cdot} - \mu, \qquad \mu_{i\cdot} = \sum_j \mu_{ij}/b,$$
$$\mu = \sum_i \mu_{ij}/ab.$$
$$\delta_{\cdot j} = \mu_{\cdot j} - \mu \qquad \mu_{\cdot j} = \sum_i \mu_{ij}/a,$$

*24. Suppose levels $A_1, \ldots, A_a$ are chosen for a factor A, levels $B_1, \ldots, B_b$ for a factor B, levels $C_1, \ldots, C_c$ for a factor C. Let y_{ijks} be the sth response at the combination $A_iB_jC_k$; $s = 1, \ldots, n$. Let $\mathbf{v}_{000}$ be the $\mathbf{1}$-vector; let $\mathbf{v}_{i00}, \mathbf{v}_{0j0}, \mathbf{v}_{00k}, \mathbf{v}_{ij0}, \mathbf{v}_{i0k}, \mathbf{v}_{0jk}, \mathbf{v}_{ijk}$ be indicator vectors for $A_i, B_j, C_k, A_iB_j, A_iC_k, B_jC_k, A_iB_jC_k$, respectively.



Problem 24

(i) Check the ordering → by "contained-in"; check the orthogonality of extensions (see the accompanying figure).

(ii) In the table check the following entries: dimension, projection. Derive two expressions for each squared length.

	Space	Dimension	Projection	Squared Length
Mean	$L(\{\mathbf{v}_{000}\})$	1	$(\bar{y}_{\cdots\cdot})$	
A	$L(\{\mathbf{v}_{i00}\})$	a	$(\bar{y}_{i\cdots})$	
B	$L(\{\mathbf{v}_{0j0}\})$	b	$(\bar{y}_{\cdot j\cdot\cdot})$	
C	$L(\{\mathbf{v}_{00k}\})$	c	$(\bar{y}_{\cdot\cdot k\cdot})$	
AB	$L(\{\mathbf{v}_{ij0}\})$	ab	$(\bar{y}_{ij\cdot\cdot})$	
BC	$L(\{\mathbf{v}_{0jk}\})$	bc	$(\bar{y}_{\cdot jk\cdot})$	
AC	$L(\{\mathbf{v}_{i0k}\})$	ac	$(\bar{y}_{i\cdot k\cdot})$	
ABC	$L(\{\mathbf{v}_{ijk}\})$	abc	$(\bar{y}_{ijk\cdot})$	
	R^N	$abcn$	(y_{ijks})	

(iii) Derive expressions for the entries in the corresponding analysis-of-variance table. This is a *three-factor factorial design*.

*25. The location and scale subgroups are examined for the regression model (Section 7) by using full matrix notation and are examined for the progression model (Section 11) by using the location-scale symbol $[\cdot, \cdot]$ of Problem 27, Chapter One. Check the details in Section 7 using the location-scale symbol, and the details in Section 11 using the full matrix notation.

*26. For the progression model derive expressions for the structural distributions, $g_L^*(\boldsymbol{\mu}: Y)d\boldsymbol{\mu}$ for $\boldsymbol{\mu}$ and $g_S^*(\mathfrak{T}: Y)d\mathfrak{T}$ for $\mathfrak{T}$ (Section 11).

*27. (*Continuation*). Derive the marginal structural distributions for $\boldsymbol{\mu}$ and $\mathfrak{T}$ for the case of normal error (Section 12).

*28. Derive the general Wishart distribution (Notes and References).

29. Consider the progression group

$$G = \left\{ g = \begin{bmatrix} c_1 & & & 0 \\ k_{21} & c_2 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ k_{p1} & \cdots & k_{p\,p-1} & c_p \end{bmatrix} : \begin{matrix} -\infty < k_{jj'} < \infty \\ 0 < c_j < \infty \end{matrix} \right\}$$

operating on points

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{p1} & \cdots & y_{pn} \end{bmatrix}$$

in Euclidean space R^{pn} by matrix multiplication.

(i) Check that G is a group and that the group is unitary on R^{pn} provided $n \geq p$ and certain trivial points are excluded.

(ii) In the pattern of Section 10 define a transformation variable and derive the invariant differential on R^{pn} and the left and right invariant differentials on G.

(iii) From properties of the invariant differentials deduce the value of the Jacobian

$$\left|\frac{\partial g^{-1}}{\partial g}\right|.$$

30 (*Continuation*). Consider *the simple progression model* for n observations on p responses:

$$f(E)\,dE,$$
$$Y = \mathcal{T}E,$$

where $\mathcal{T}$ is an element of the group G.

(i) Derive the distribution of the error position variable $[E]$ given orbit.

(ii) Derive the structural distribution for $\mathcal{T}$.

31 (*Continuation*). For the case of standard normal errors derive:

(i) The distribution of error position $[E]$ given orbit.

(ii) The distribution of the error inner product matrix $S(E) = EE' = [E][E]'$ given orbit.

(iii) The structural distribution for $\mathcal{T}$.

(iv) The structural distribution for $\Sigma = \mathcal{T}\mathcal{T}'$.

***32.** *Regression progression model.* Consider an error variable E:

$$E = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \\ e_{11} & \cdots & e_{1n} \\ \vdots & & \vdots \\ e_{p1} & \cdots & e_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1' \\ \vdots \\ \mathbf{v}_r' \\ \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{bmatrix} = \begin{bmatrix} V \\ \underline{E} \end{bmatrix}$$

with error distribution

$$f(E)\,dE = f(\underline{E})\,d\underline{E} = \prod_{i=1}^{n} f(e_{1i}, \ldots, e_{pi}) \prod_{i=1}^{n} (de_{1i} \cdots de_{pi}).$$

Consider a quantity

$$\theta = \begin{bmatrix} 1 & & 0 & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ 0 & & 1 & 0 & \cdots & 0 \\ \beta_{11} & \cdots & \beta_{1r} & \sigma_{(1)} & & 0 \\ \beta_{21} & \cdots & \beta_{2r} & \tau_{21} & \sigma_{(2)} & \\ \vdots & & \vdots & \vdots & & \ddots \\ \beta_{p1} & \cdots & \beta_{pr} & \tau_{p1} & \cdots\ \tau_{pp-1} & \sigma_{(p)} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \mathcal{B} & \mathcal{T} \end{bmatrix},$$

or, equivalently (general location-scale notation in Problem 27, Chapter One),

$$\underline{\theta} = [\mathcal{B}, \mathcal{T}].$$

And consider a response matrix Y:

$$Y = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \\ y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{p1} & \cdots & y_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1' \\ \vdots \\ \mathbf{v}_r' \\ \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_p' \end{bmatrix} = \begin{bmatrix} V \\ \underline{Y} \end{bmatrix}.$$

The *regression-progression model* is

$$\begin{matrix} f(E)\,dE, \\ Y = \theta E, \end{matrix} \quad \text{or} \quad \begin{matrix} f(\underline{E})\,d\underline{E}, \\ \underline{Y} = \mathcal{B}V + \mathcal{T}\underline{E}. \end{matrix}$$

(i) Check the equivalence of the two kinds of notation: full matrix; location-scale with matrix arguments.

(ii) Consider the regression-progression group:

$$G = \left\{ g = \begin{bmatrix} I & 0 \\ B & T \end{bmatrix} : \begin{matrix} B \text{ is } p \times r \text{ matrix} \\ T \text{ is } p \times p \text{ positive-lower-triangular} \end{matrix} \right\}.$$

Check that G is a group. Describe the orbits on R^{pn} using the L^+ notation; show that G is unitary on R^{pn} if $n \geq r + p$ and a certain degenerate set of points is deleted.

***33** (*Continuation*). Define a variable $[Y]$:

$$[Y] = \begin{bmatrix} I & 0 \\ B(Y) & T(Y) \end{bmatrix}$$

and a point $D(Y)$ in R^{pn}:

$$D(Y) = \begin{bmatrix} \mathbf{v}_1' \\ \vdots \\ \mathbf{v}_r' \\ \mathbf{d}_1' \\ \vdots \\ \mathbf{d}_p' \end{bmatrix} = \begin{bmatrix} V \\ \underline{D}(Y) \end{bmatrix}.$$

Use the regression coefficients of Section 3 and the natural extension of the regression coefficients and residual unit vectors in Section 10. Show that $[Y]$ is a transformation variable and $D(Y)$ is a reference point; check the alternative notation: $[B(Y), T(Y)]$ and $\underline{D}(Y)$.

*34 (*Continuation*)). Verify the following invariant differentials:

$$dm(Y) = \frac{dY}{|[Y]|^n},$$

$$d\mu(g) = \frac{dg}{|g|_\Delta} = \frac{dB\,dT}{|T|^r\,|T|_\Delta},$$

$$d\nu(g) = \frac{dg}{|g|_\nabla} = \frac{dB\,dT}{|T|_\nabla},$$

$$\Delta(g) = \frac{|g|_\nabla}{|g|_\Delta} = \frac{|T|_\nabla}{|T|^r\,|T|_\Delta}.$$

*35 (*Continuation*). Derive the following distributions:

$$g([E]:D)\,d[E] = k(D)f([E]D)\,|[E]|^n \frac{d[E]}{|[E]|_\Delta}$$

$$= k(D)\prod_{i=1}^{n} f\left(B\begin{bmatrix} v_{1i} \\ \cdot \\ \cdot \\ \cdot \\ v_{ri}\end{bmatrix} + T\begin{bmatrix} d_{1i} \\ \cdot \\ \cdot \\ \cdot \\ d_{pi}\end{bmatrix}\right) s_{(1)}^{n-(r+1)} \cdots s_{(p)}^{n-(r+p)}\,dB\,dT.$$

$$g^*(\theta : Y)\,d\theta = k(D)f(\theta^{-1}Y)\frac{|[Y]|^n}{|\theta|^n}\frac{|[Y]|_\nabla}{|[Y]|_\Delta}\,d\nu(\theta)$$

$$= k(D)\prod_{i=1}^{n} f\left(\mathcal{T}^{-1}\left(\begin{bmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi}\end{bmatrix} - \mathcal{B}\begin{bmatrix} v_{1i} \\ \cdot \\ \cdot \\ \cdot \\ v_{ri}\end{bmatrix}\right)\right)\left(\frac{s_{(1)}(Y)\cdots s_{(p)}(Y)}{\sigma_{(1)}\cdots\sigma_{(p)}}\right)^n.$$

$$\frac{s_{(1)}^{p}(Y)\cdots s_{(p)}^{1}(Y)}{s_{(1)}^{r+1}(Y)\cdots s_{(p)}^{r+p}(Y)}\qquad\frac{d\mathcal{B}\,d\mathcal{T}}{\sigma_{(1)}^{p}\cdots\sigma_{(p)}^{1}}$$

*36 (*Continuation*). Derive the location distribution $g_L(H:D)\,dH$ for the error variable† $H = T^{-1}B$; derive the structural distribution $g_L^*(\mathcal{B}:Y)\,d\mathcal{B}$ for $\mathcal{B}$. Derive the scale distribution $g_S(T:D)\,dT$ for the error variable $T = T(E)$. Derive the structural distribution $g_S^*(\mathcal{T}:Y)\,d\mathcal{T}$ for $\mathcal{T}$.

† *Note.* H is a $p \times r$ matrix, an analog of the vector $\mathbf{t}$ in the regression and progression models; the capital T has been used already for the positive-lower-triangular scale matrix.

*37. (*Continuation*). For the case of normal error

$$f(\underline{E})\,d\underline{E} = (2\pi)^{-np/2}\exp\{-\tfrac{1}{2}\operatorname{tr}\underline{E}\,\underline{E}'\}\,d\underline{E},$$

determine the form of the distributions in Problem 35 and structural distributions in Problem 36.

*38 (*Continuation*). Derive the standard Wishart distribution and the general Wishart distribution for the regression-progression model. Note the similarity in form to the distribution for the progression model case—a change of degrees of freedom.

39. Consider a matrix V of structural vectors and a matrix $\underline{Y}$ of corresponding response vectors; let

$$Y = \begin{bmatrix} V \\ \hline \underline{Y}\end{bmatrix}\qquad YY' = \begin{bmatrix} V \\ \hline \underline{Y}\end{bmatrix}\begin{bmatrix} V \\ \hline \underline{Y}\end{bmatrix}' = \left[\begin{array}{c|c} S_{11} & S_{12} \\ \hline S_{21} & S_{22}\end{array}\right].$$

(i) Show that the regression coefficients of $\underline{Y}$ on V are $S_{21}S_{11}^{-1}$: coefficients for the first response are in the first row, . . . , coefficients for the last response are in the last row.

(ii) Show that the matrix of residual vectors is $\underline{Y} - S_{21}S_{11}^{-1}V$.

(iii) Show that the inner-product matrix of residuals is $S_{22} - S_{21}S_{11}^{-1}S_{12}$.

(iv) Consider the inner-product matrix for a general residual $\underline{Y} - BV$:

$$(\underline{Y} - BV)(\underline{Y} - BV)'.$$

Complete the quadratic form in B; and thereby show that $B = S_{21}S_{11}^{-1}$ gives a *minimum* inner-product matrix of residuals.

40. (i) Determine W to satisfy the $(r + p)$-partitioned-matrix multiplication,

$$\begin{bmatrix} I & 0 \\ W & I\end{bmatrix}\begin{bmatrix} A & C \\ B & D\end{bmatrix} = \begin{bmatrix} A & C \\ 0 & D - BA^{-1}C\end{bmatrix},$$

and show that

$$\begin{vmatrix} A & C \\ B & D\end{vmatrix} = |A|\,|D - BA^{-1}C|.$$

(ii) Show that

$$|I - K_1K_2| = |I - K_2K_1|,$$

where K_1 is $p \times r$, K_2 is $r \times p$, and I in each case is an appropriate identity matrix.

PART II

Methods of Analysis

CHAPTER FOUR

Conditional Analysis

The structural model describes a certain kind of process or system: the form of the error distribution is known from theory or from experience with the same or related systems; the physical quantity is a transformation from a group, the transformation that carries an error value into a response value. The presence of the group is important: it permits the identification of characteristics of an error value; and with multiple observations the group allows the form of the error distribution to be determined; it allows the observer to see *into* the system.

Not every physical quantity can be identified as a transformation from a group. This chapter examines a variety of models in which *part* of the physical quantity is a transformation from a group. Chapters Seven and Eight examine models in which *no part* of the quantity is a transformation from a group.

With a structural model a change in the quantity is a change in a transformation that applies to the error value; it produces a corresponding change in the response. Without this structuring relationship linking the quantity and the response, there remains only the frequency distribution that describes possible response values for each value for the quantity—*the classical model of statistics*. For the models in this chapter the transformation part of the quantity can be analyzed in the framework of a structural model, and the remaining part can be analyzed by methods appropriate to the classical model.

1 PROBABILITY AND LIKELIHOOD FUNCTIONS

Consider a model with a response variable x taking possible values in a space $\mathfrak{X}$ and with a quantity θ taking possible values in a space Ω. Suppose there is no structuring relationship by which a change in θ can be related to a change in x; specifically, suppose there is the minimum for a statistical model—a frequency distribution $f(x:\theta)$ for the response variable x for each

value for θ, *the classical model of statistics*. Suppose also that the space $\mathfrak{X}$ has a countable number of points with $f(x:\theta)$ as the probability function, or, alternatively, that the space $\mathfrak{X}$ is an open subset of a Euclidean space with $f(x:\theta)$ as the probability density function relative to Euclidean volume. *The statistical model then has a frequency distribution $f(x:\theta)$ for x in $\mathfrak{X}$ with θ in Ω, and it has an observed value*† *x_0 for the response variable x.*

Consider first the case of a countable space $\mathfrak{X}$. Within the model an observed x_0 has its *labeling* x_0 and its probability of occurrence $f(x_0:\theta)$ as *a function of the possible values θ for the quantity*. The labeling x_0 is irrelevant because of the assumed absence of any structure relating the quantity to the response. Within the model, then, an observed value has *only the function $f(x_0:\theta)$ of θ* as its essential identification; two points x_0', x_0'' with the same function $f(x_0':\theta) = f(x_0'':\theta)$ are not distinguished. See Figure 1. The *reduced* statistical model then has a frequency distribution $f(x:\theta)$ and it has a realized function $f(x_0:\theta)$ giving probability of occurrence as a function of θ. Some inference methods, particularly for a large number of observations, are examined in Chapter Seven. A basic principle uses the function $f(x_0:\theta)$ to assess the various possible values for θ, perhaps choosing as a single preferred value the value $\hat{\theta}$ that maximizes the function $f(x_0:\theta)$.

Now consider the case of an open set $\mathfrak{X}$ in Euclidean space. Within the model an observed x_0 has only its labeling x_0 and its probability of occurrence $f(x_0:\theta)\,dx_0$ (in an element dx_0 that includes x_0). Again the labeling is irrelevant because of the assumed absence of any structure relating the quantity to the response. Within the model, then, an observed value x_0 has *only the function* $f(x_0:\theta)\,dx_0$ (with dx_0 unspecified in magnitude) as its essential identification. This can be expressed more compactly by defining the *likelihood function from the observed x_0*:

$$L(x_0:\theta) = \{kf(x_0:\theta): \quad 0 < k < \infty\}.$$

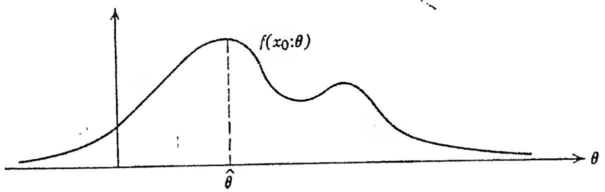


Figure 1 The probability function for an observed value x_0.

† In most analyses a single letter can be used to designate both a variable and a corresponding realized value, the distinction being made by the context. It is convenient, however, in the present context to make the distinction explicit and use a subscript o to distinguish an observed or realized value.

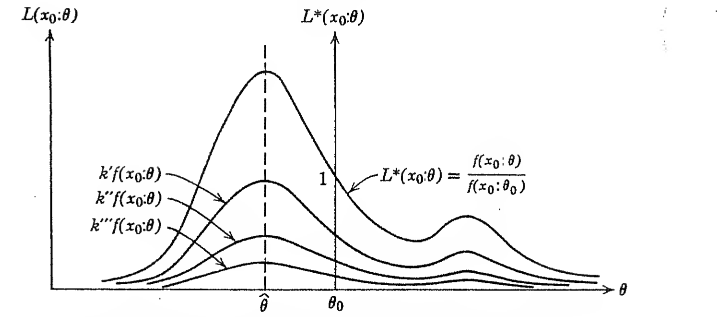


Figure 2 Representative elements $k'f(x_0:\theta)$, $k''f(x_0:\theta)$, $k'''f(x_0:\theta)$, of the likelihood function $L(x_0:\theta)$. The likelihood ratio $L^*(x_0:\theta)$ relative to θ_0.

The likelihood function, in fact, is a *set* of functions of θ, all of the same form and differing only in the positive *multiplicative* constant k. This definition accommodates the unspecified element dx_0. See Figure 2. A second, more formal expression for the likelihood function is

$$L(x_0:\theta) = R^+(x_0)f(x_0:\theta),$$

where $R^+(x)$ is the map that carries any *point* x in $\mathfrak{X}$ to the *set* $R^+ = (0, \infty)$ of positive real numbers. Note that $kR^+(x_0) = R^+$ for any positive number k. If $f(x_0:\theta) \neq 0$ for each θ then the unspecified constant can be avoided by using a *likelihood ratio* relative to some reference value θ_0:

$$L^*(x_0:\theta) = \frac{f(x_0:\theta)}{f(x_0:\theta_0)}.$$

The likelihood function can be expressed alternatively as the *log-likelihood function from the observed x_0*:

$$l(x_0:\theta) = \{c + \ln f(x_0:\theta): \quad -\infty < c < \infty\} = R(x_0) + \ln f(x_0:\theta),$$

where $R(x)$ is the map that carries any *point* x in $\mathfrak{X}$ to the *set* $R = (-\infty, \infty)$ of real numbers. The log-likelihood function, in fact, is a set of functions of θ all of the same form and differing only in the *additive* constant c. See Figure 3. The value $-\infty$ must be allowed for $\ln f(x_0:\theta)$ to correspond to the value 0 for $f(x_0:\theta)$.

Consider further the case of an open set $\mathfrak{X}$ in Euclidean space. Within the model an observed value x_0 has only the *likelihood function* $L(x_0:\theta)$ as its essential identification; the likelihood function gives the *relative* probability of occurrence of x_0 under various possible values for θ; two points x_0', x_0'' with the same likelihood function are not distinguished. The *reduced* statistical model then has a frequency function $f(x:\theta)$ and it has a realized likelihood

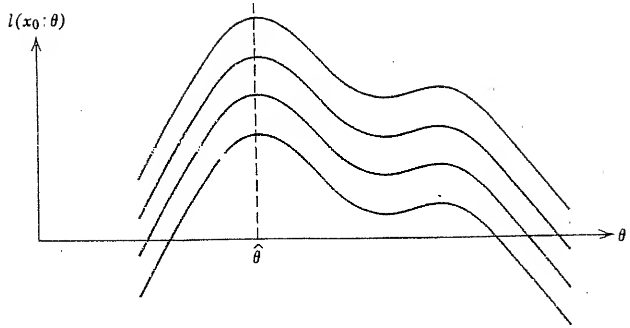


Figure 3 Representative elements $c' + \ln f(x_0:\theta)$, $c'' + \ln f(x_0:\theta)$, $c''' + \ln f(x_0:\theta)$, $c'''' + \ln f(x_0:\theta)$ for the log-likelihood function $l(x_0:\theta)$.

function $L(x_0:\theta)$ giving relative probability of occurrence as a function of θ. Some inference methods, particularly for a large number of observations, are examined in Chapter Eight. A basic principle uses the function $L(x_0:\theta)$ to assess, one-with-respect-to-another, the various values for θ, perhaps choosing as a single preferred value the value $\hat{\theta}$ that maximizes the likelihood function $L(x_0:\theta)$.

2 A CONDITIONAL MODEL AND MARGINAL LIKELIHOOD

Consider now a model that is partly structural and partly classical. Let E be an error variable on an open set $\mathfrak{X}$ in a Euclidean space R^N. Let θ be a *primary quantity*, an element of a unitary group G of transformations of $\mathfrak{X}$ onto $\mathfrak{X}$ (with Assumption 3 in Chapter Two). Let X be an observed response that is produced by the transformation θ applied to a realized error value E. And suppose that the error distribution

$$f(E:\lambda)\,dE$$

is known except for an *additional quantity* λ. This gives the

Conditional Structural Model

$$f(E:\lambda)\,dE,$$
$$X = \theta E,$$

with additional quantity λ. The model has an error variable E with distribution dependent on the quantity λ; and it has a structural equation in which a realized error value E is transformed by the quantity θ to give the response X. If the additional quantity λ is known in value, then the conditional structural model is an ordinary structural model.

Now suppose there is no outside information concerning θ and λ. For an assumed value for the quantity λ the structural model produces a reduced structural model

$$g_\lambda([E]:D(X))\,d[E],$$
$$[X] = \theta[E],$$

where

$$g_\lambda([E]:D)\,d[E] = k_\lambda(D)f([E]D:\lambda)J_N(E)\,d\mu([E]).$$

The corresponding structural distribution for θ conditional on λ is

$$g_\lambda^*(\theta:X)\,d\theta = k_\lambda(D)f(\theta^{-1}X:\lambda)J_N(\theta^{-1}X)\,\Delta([X])\,d\nu(\theta).$$

For an assumed value for λ this distribution is the basis for inference concerning θ.

Now consider inference concerning λ. The structural equation gives *no* information concerning the error position $[E]$, but it gives the value of the error orbit $GE = GX$. The distribution that describes the origin of the error position $[E]$ is a distribution that involves λ; it can give no information concerning λ without the realized position $[E]$. This leaves only the known orbit $GE = GX$ and the distribution that describes its origin. The distribution that describes the origin of the orbit $GE = GX$ involves the quantity λ; it is a classical model. The likelihood function for the known orbit based on this classical model is now derived and is called the marginal likelihood function for λ.

The probability element for E based on Euclidean volume is

$$f(E:\lambda)\,dE.$$

The conditional probability element for $[E]$ given the orbit $D(E) = D$ is

$$k_\lambda(D)f([E]D:\lambda)\frac{J_N(E)}{J_L([E])}\,d[E].$$

The marginal probability element for the orbit $D = D(E)$ can be obtained by dividing the full element by the conditional element

$$\frac{1}{k_\lambda(D)}\,\frac{J_L([E])}{J_N(E)}\,\frac{dE}{d[E]}.$$

The marginal element based on differentials at the point X rather than at E on the orbit D is

$$\frac{1}{k_\lambda(D)}\,\frac{J_L([X])}{J_N(X)}\,\frac{dX}{d[X]}.$$

The likelihood function based on D is

$$L(D:\lambda) = R^+(D)\frac{1}{k_\lambda(D)};$$

the omitted factors do not involve λ and have been incorporated into $R^+(D)$. This is the *marginal likelihood function for* λ. This marginal likelihood function is the basis for inference concerning the quantity λ.

3 THE MEASUREMENT MODEL WITH AN ERROR QUANTITY

Consider the measurement model of Chapter One and suppose that the error distribution

$$f(e:\beta)\, de$$

involves a *shape* or *form* quantity β. As examples consider

$$f_1(e:\beta) = k \exp\{-\tfrac{1}{2}|e|^\beta\}, \qquad \beta > 0,$$

$$f_2(e:\beta) = k(1 + \exp\{-\beta e\})^{-1} \exp\left\{-\frac{e^2}{2}\right\}.$$

For n measurements the model is

$$\prod_1^n f(e_i:\beta) \prod_1^n de_i,$$

$$\mathbf{x} = [\mu, \sigma]\mathbf{e}.$$

This is a conditional structural model with additional quantity β.

For an assumed value for the quantity β the reduced structural model is

$$k_\beta(\mathbf{d}) \prod_1^n f(\bar{e} + s_e d_i:\beta) s_e^n \frac{d\bar{e}\, ds_e}{s_e^2},$$

$$[\bar{x}, s_x] = [\mu, \sigma][\bar{e}, s_e].$$

The structural distribution for $[\mu, \sigma]$ conditional on β (Section 18, Chapter One) is

$$k_\beta(\mathbf{d}) \prod_1^n f\left(\frac{x_i - \mu}{\sigma}:\beta\right)\left(\frac{s_x}{\sigma}\right)^n \frac{1}{s_x}\frac{d\mu\, d\sigma}{\sigma}.$$

The probability element for $\mathbf{e}$ based on Euclidean volume is

$$\prod_1^n f(e_i:\beta) \cdot \prod_1^n de_i.$$

The conditional probability element for $[\bar{e}, s_e]$ given the orbit $\mathbf{d}$ is

$$k_\beta(\mathbf{d}) \prod_1^n f(\bar{e} + s_e d_i:\beta) \frac{s_e^n}{s_e^2}\, d\bar{e}\, ds_e;$$

it is expressed in terms of Euclidean area on the positive affine group. The vector $[\bar{e}, s_e]\mathbf{d}$ is a point on the orbit through $\mathbf{d}$ in Euclidean space R^n; let

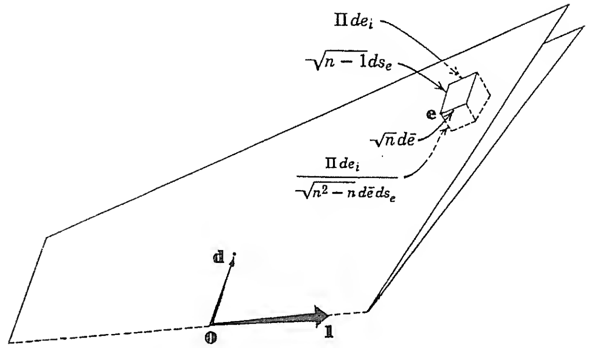


Figure 4. Euclidean measure on the orbit: $\sqrt{n}d\bar{e}\sqrt{n-1}\, ds_e$. Euclidean volume measure cross-sectional to the orbit: $\Pi de_i/\sqrt{n^2 - n}\, d\bar{e}\, ds_e$.

$d[\bar{e}, s_e]\mathbf{d}$ designate Euclidean area on this orbit. The ratio of Euclidean area on the orbit to Euclidean area on the group is

$$K([\bar{e}, s_e]:\mathbf{d}) = \left|\frac{\partial[\bar{e}, s_e]\mathbf{d}}{\partial[\bar{e}, s_e]}\right| = \sqrt{n}\sqrt{n-1};$$

see the example at the end of Section 3, Chapter Two; also Figure 15, Chapter One. The conditional probability element for $[\bar{e}, s_e]$ can then be written

$$k_\beta(\mathbf{d}) \prod_1^n f(\bar{e} + s_e d_i:\beta) \frac{s_e^n}{s_e^2\sqrt{n^2 - n}}\, d[\bar{e}, s_e]\mathbf{d}$$

as based on Euclidean area on the orbit; see Figure 4.

The marginal probability element for the orbit $\mathbf{d} = \mathbf{d}(\mathbf{e})$ is obtained by division:

$$\frac{1}{k_\beta(\mathbf{d})} \cdot \frac{\sqrt{n^2 - n}}{s_e^{n-2}} \cdot \frac{\prod de_i}{d[\bar{e}, s_e]\mathbf{d}};$$

it is expressed in terms of $(n - 2)$-dimensional Euclidean volume cross-sectional to the orbit. The marginal probability element for the orbit as based on $(n - 2)$-dimensional Euclidean volume dv *at the observation vector* $\mathbf{x}$ is

$$\frac{1}{k_\beta(\mathbf{d})} \cdot \frac{\sqrt{n^2 - n}}{s_x^{n-2}} \cdot dv.$$

The *marginal likelihood function for* β *is*

$$L(\mathbf{d}:\beta) = R^+(\mathbf{d}) \frac{1}{k_\beta(\mathbf{d})}.$$

The structural distribution for $[\mu, \sigma]$ provides the basis for inference concerning $[\mu, \sigma]$ for an assumed value for β: the marginal likelihood function provides the basis for inference concerning β.

4 A COMPOSITE RESPONSE MODEL

Consider a sequence of p response variables $y_1, \ldots, y_p$ and suppose that the distribution

$$f(e_1, \ldots, e_p : \beta)$$

of the related error sequence $e_1, \ldots, e_p$ has been identified, except for a quantity β describing form, shape, linking, or other characteristic or combination of characteristics. Let μ_j be the general level of the jth response and σ_j be the error scaling for the jth response. For n observations on the *composite* response the following model is obtained:

$$\prod f(e_{1i}, \ldots, e_{pi} : \beta) \prod de_{ji},$$

$$\mathbf{y}_1 = [\mu_1, \sigma_1]\mathbf{e}_1$$

$$\vdots$$

$$\mathbf{y}_p = [\mu_p, \sigma_p]\mathbf{e}_p.$$

The model is a conditional structural model ($n \geq 2$) with additional quantity β.

An example of a bivariate error distribution with additional quantity ρ is

$$f(e_1, e_2 : \rho) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left\{ -\frac{e_1^2 - 2\rho e_1 e_2 + e_2^2}{2(1 - \rho^2)} \right\},$$

the *bivariate normal error distribution with correlation* ρ.

The quantity $[\mu_j, \sigma_j]$ belongs to the positive affine group

$$G_j = \left\{ [a_j, c_j] : \begin{matrix} -\infty < a_j < \infty \\ 0 < c_j < \infty \end{matrix} \right\}$$

on the Euclidean space R^n of the jth response vector. The composite quantity $([\mu_1, \sigma_1], \ldots, [\mu_p, \sigma_p])$ belongs to a group

$$G = \{([a_1, c_1], \ldots, [a_p, c_p]) : \quad [a_j, c_j] \in G_j\},$$

a *direct product of the groups* $G_1, \ldots, G_p$ (actually the pth power of the positive affine group since $G_1, \ldots, G_p$ are equivalent); the multiplication is coordinate-by-coordinate:

$$([a_1, c_1], \ldots, [a_p, c_p]) \cdot ([A_1, C_1], \ldots, [A_p, C_p])$$
$$= ([a_1, c_1][A_1, C_1], \ldots, [a_p, c_p][A_p, C_p]).$$

The invariant differentials and modular function can be obtained by combining those for the component groups.

For an assumed value for the quantity β the reduced structural model is

$$k_\beta(\mathbf{d}_1, \ldots, \mathbf{d}_p) \prod f(\bar{e}_1 + s_{e_1} d_{1i}, \ldots, \bar{e}_p + s_{e_p} d_{pi} : \beta) \cdot (s_{e_1} \cdots s_{e_p})^n \prod \frac{d\bar{e}_j \, ds_{e_j}}{s_{e_j}^2},$$

$$[\bar{y}_1, s_{y_1}] = [\mu_1, \sigma_1][\bar{e}_1, s_{e_1}],$$

$$\vdots$$

$$[\bar{y}_p, s_{y_p}] = [\mu_p, \sigma_p][\bar{e}_p, s_{e_p}].$$

The structural distribution for the composite quantity conditional on β is

$$k_\beta(\mathbf{d}_1, \ldots, \mathbf{d}_p) \prod f\left(\frac{y_{1i} - \mu_1}{\sigma_1}, \ldots, \frac{y_{pi} - \mu_p}{\sigma_p} : \beta\right) \cdot \frac{(s_{y_1} \cdots s_{y_p})^{n-1}}{(\sigma_1 \cdots \sigma_p)^n} \prod \frac{d\mu_j \, d\sigma_j}{\sigma_j}.$$

The marginal probability element for the orbit

$$(\mathbf{d}_1, \ldots, \mathbf{d}_p) = (\mathbf{d}_1(\mathbf{e}_1), \ldots, \mathbf{d}_p(\mathbf{e}_p))$$

can be obtained by the methods in the preceding section applied to each response vector:

$$\frac{1}{k_\beta(\mathbf{d}_1, \ldots, \mathbf{d}_p)} \cdot \frac{(n^2 - n)^{p/2}}{(s_{y_1} \cdots s_{y_p})^{n-2}} \cdot dv;$$

the element dv measures $(n - 2)p$-dimensional Euclidean volume cross-sectional to the orbit at the observed composite response. The marginal likelihood function for β is

$$L(\mathbf{d}_1, \ldots, \mathbf{d}_p : \beta) = R^+(\mathbf{d}_1, \ldots, \mathbf{d}_p) \frac{1}{k_\beta(\mathbf{d}_1, \ldots, \mathbf{d}_p)}.$$

For an example consider the bivariate normal error distribution with correlation ρ. The normalizing constant for the conditional error distribution

can be obtained by integration:

$$k_\beta^{-1}(\mathbf{d}_1, \mathbf{d}_2)$$

$$= \frac{1}{(2\pi)^n(1-\rho^2)^{n/2}}$$

$$\cdot \int \exp\left\{-\sum \frac{(\bar{e}_1 + s_{e_1}d_{1i})^2 - 2\rho(\bar{e}_1 + s_{e_1}d_{1i})\cdot(\bar{e}_2 + s_{e_2}d_{2i}) + (\bar{e}_2 + s_{e_2}d_{2i})^2}{2(1-\rho^2)}\right\}$$

$$\cdot (s_{e_1}s_{e_2})^{n-2}\, d\bar{e}_1\, d\bar{e}_2\, ds_{e_1}\, ds_{e_2}$$

$$= \int \frac{1}{(2\pi)(1-\rho^2)^{1/2}} \exp\left\{-n\frac{\bar{e}_1^2 - 2\rho\bar{e}_1\bar{e}_2 + \bar{e}_2^2}{2(1-\rho^2)}\right\} d\sqrt{n}\,\bar{e}_1\, d\sqrt{n}\,\bar{e}_2$$

$$\cdot \frac{1}{n(2\pi)^{n-1}(1-\rho^2)^{(n-1)/2}} \int \exp\left\{-(n-1)\frac{s_{e_1}^2 - 2\rho r s_{e_1}s_{e_2} + s_{e_2}^2}{2(1-\rho^2)}\right\}$$

$$\cdot (s_{e_1}s_{e_2})^{n-2}\, ds_{e_1}\, ds_{e_2}$$

$$= \frac{2^{n-1}(1-\rho^2)^{(n-1)/2}}{(2\pi)^{n-1}n(n-1)^{n-1}} \int_0^\infty\!\!\int \exp\{-t_1^2 - t_2^2 + 2\rho r t_1 t_2\}(t_1t_2)^{n-2}\, dt_1\, dt_2$$

$$= \frac{2^{n-1}(1-\rho^2)^{(n-1)/2}}{(2\pi)^{n-1}n(n-1)^{n-1}} \sum_{\alpha=0}^{\infty} \frac{(2\rho r)^\alpha}{\alpha!} \int_0^\infty\!\!\int \exp\{-t_1^2 - t_2^2\}(t_1t_2)^{n-2+\alpha}\, dt_1\, dt_2$$

$$= \frac{2^{n-1}(1-\rho^2)^{(n-1)/2}}{(2\pi)^{n-1}n(n-1)^{n-1}} \sum_{\alpha=0}^{\infty} \frac{(2\rho r)^\alpha}{\alpha!}\frac{1}{2^2} \int_0^\infty\!\!\int \exp\{-t_1^2 - t_2^2\}(t_1^2t_2^2)^{(n-1+\alpha)/2-1}\, dt_1^2\, dt_2^2$$

$$= \frac{2^{n-3}(1-\rho^2)^{(n-1)/2}}{(2\pi)^{n-1}n(n-1)^{n-1}} \sum_{\alpha=0}^{\infty} \frac{(2\rho r)^\alpha}{\alpha!} \Gamma^2\left(\frac{n-1+\alpha}{2}\right)$$

$$= \frac{2^{n-3}(1-\rho^2)^{(n-1)/2}}{(2\pi)^{n-1}n(n-1)^{n-1}} H_{n-1}(\rho r),$$

where

$$H_n(t) = \sum_{\alpha=0}^{\infty} \frac{(2t)^\alpha}{\alpha!}\Gamma^2\left(\frac{n+\alpha}{2}\right).$$

The preceding simplification involves a number of steps. In the first step the error density function is substituted. In the second step the terms in the exponential are expanded and rearranged,

$$\sum_i d_{ji} = 0, \qquad \sum_i d_{ji}^2 = n-1, \qquad \sum_i d_{1i}\,d_{2i} = (n-1)r,$$

and expressed in terms of the sample correlation coefficient r,

$$r = \frac{(n-1)^{-1}\sum(y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{s_{y_1}s_{y_2}}$$

$$= \frac{(n-1)^{-1}\sum(e_{1i} - \bar{e}_1)(e_{2i} - \bar{e}_2)}{s_{e_1}s_{e_2}} = \frac{1}{n-1}\sum d_{1i}\,d_{2i},$$

between $\mathbf{y}_1$ and $\mathbf{y}_2$, between $\mathbf{e}_1$ and $\mathbf{e}_2$, or between $\mathbf{d}_1$ and $\mathbf{d}_2$; the first part of the expression, an integral, has value 1. In the third step the substitution

$$t_j = \frac{\sqrt{n-1}\, s_{e_j}}{\sqrt{2}\sqrt{1-\rho^2}}$$

is made. In the remaining steps the cross-term in the exponential is expanded in a series and is integrated term by term.

The conditional error distribution given the orbit then has the form

$$\frac{k_\beta(\mathbf{d}_1, \mathbf{d}_2)}{(2\pi)^n(1-\rho^2)^{n/2}}$$

$$\cdot \exp\left\{-\sum \frac{(\bar{e}_1 + s_{e_1}d_{1i})^2 - 2\rho(\bar{e}_1 + s_{e_1}d_{1i})\cdot(\bar{e}_2 + s_{e_2}d_{2i}) + (\bar{e}_2 + s_{e_2}d_{2i})^2}{2(1-\rho^2)}\right\}$$

$$\cdot (s_{e_1}s_{e_2})^{n-2}\, d\bar{e}_1\, d\bar{e}_2\, ds_{e_1}\, ds_{e_2}$$

$$= \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-n\frac{\bar{e}_1^2 - 2\rho\bar{e}_1\bar{e}_2 + \bar{e}_2^2}{2(1-\rho^2)}\right\} d\sqrt{n}\,\bar{e}_1\, d\sqrt{n}\,\bar{e}_2$$

$$\cdot \frac{(n-1)^{n-1}}{2^{n-3}(1-\rho^2)^{n-1}H_{n-1}(\rho r)} \exp\left\{-(n-1)\frac{s_{e_1}^2 - 2\rho r s_{e_1}s_{e_2} + s_{e_2}^2}{2(1-\rho^2)}\right\}$$

$$\cdot (s_{e_1}s_{e_2})^{n-2}\, ds_{e_1}\, ds_{e_2}.$$

And the marginal probability element for the orbit $(\mathbf{d}_1, \mathbf{d}_2)$ at the observed response is

$$\frac{1}{k_\beta(\mathbf{d}_1, \mathbf{d}_2)} \cdot \frac{n^2 - n}{(s_{y_1}s_{y_2})^{n-2}} \cdot dv = \frac{2^{n-3}(1-\rho^2)^{(n-1)/2}H_{n-1}(\rho r)}{(2\pi)^{n-1}(n-1)^{n-2}(s_{y_1}s_{y_2})^{n-2}}\, dv.$$

The marginal likelihood for ρ is

$$L(\mathbf{d}_1, \mathbf{d}_2 : \rho) = R^+(\mathbf{d}_1, \mathbf{d}_2)(1-\rho^2)^{(n-1)/2}H_{n-1}(\rho r);$$

This can be expressed as a ratio relative to $\rho = 0$:

$$L^*(\mathbf{d}_1, \mathbf{d}_2 : \rho) = \frac{(1-\rho^2)^{(n-1)/2}H_{n-1}(\rho r)}{H_{n-1}(0)} = \frac{(1-\rho^2)^{(n-1)/2}H_{n-1}(\rho r)}{\Gamma^2((n-1)/2)}.$$

The marginal likelihood function $L^*(\mathbf{d}_1, \mathbf{d}_2 : \rho)$ depends on the orbit $(\mathbf{d}_1, \mathbf{d}_2)$, but it is seen to depend on the orbit *only in terms of the correlation coefficient r*. The probability element for the orbit $(\mathbf{d}_1, \mathbf{d}_2)$ for general ρ can be related to the element for $\rho = 0$:

$$h(\mathbf{d}_1, \mathbf{d}_2 : \rho)\delta(\mathbf{d}_1, \mathbf{d}_2) = L^*(\mathbf{d}_1, \mathbf{d}_2 : \rho) h(\mathbf{d}_1, \mathbf{d}_2 : 0)\delta(\mathbf{d}_1, \mathbf{d}_2).$$

The distribution for $(\mathbf{d}_1, \mathbf{d}_2)$ has its dependence on ρ isolated in the factor L^* and this factor depends on the orbit $(\mathbf{d}_1, \mathbf{d}_2)$ *only* in terms of the correlation r. An integration to obtain the marginal distribution of $r = (n-1)^{-1}(\mathbf{d}_1, \mathbf{d}_2)$ is an integration over variation in $\mathbf{d}_1, \mathbf{d}_2$ for fixed r. In this integration L^* is a constant factor; hence, if

$$h(r:0)\,dr$$

is the marginal distribution for r with $\rho = 0$, then

$$h(r:\rho)\,dr = \frac{(1-\rho^2)^{(n-1)/2} H_{n-1}(\rho r)}{\Gamma^2((n-1)/2)}\, h(r:0)\,dr$$

is the marginal distribution of r for general correlation ρ. The general distribution is obtained by *likelihood modulation* of the special distribution.

For $\rho = 0$ the distribution of r has element

$$\frac{\Gamma((n-1)/2)}{\Gamma(\frac{1}{2})\Gamma((n-2)/2)}(1-r^2)^{(n-4)/2}\,dr = \frac{2^{n-3}\Gamma^2((n-1)/2)}{\pi\Gamma(n-2)}(1-r^2)^{(n-4)/2}\,dr, \qquad -1 < r < 1;$$

note that $\Gamma(2p) = 2^{2p-1}\Gamma(p)\Gamma(p+\frac{1}{2})/\Gamma(\frac{1}{2})$. This distribution can be established by using the normal regression theory in Chapter Three to show that

$$t = \frac{\sqrt{n-2}\, r}{\sqrt{1-r^2}}$$

has a t-distribution on $n-2$ degrees of freedom conditionally given $\mathbf{e}_1$, hence marginally.

The general distribution for the correlation coefficient r is then

$$\frac{2^{n-3}}{\pi\Gamma(n-2)}(1-\rho^2)^{(n-1)/2} H_{n-1}(\rho r)(1-r^2)^{(n-4)/2}\,dr, \qquad -1 < r < 1.$$

5 THE MEASUREMENT MODEL ON THE CIRCLE

Consider a surveyor measuring a direction in the horizontal plane, or a physicist measuring a directional property on a plane surface, or an oceanographer measuring the direction of a wave train. These are applications for the measurement model on the circle.

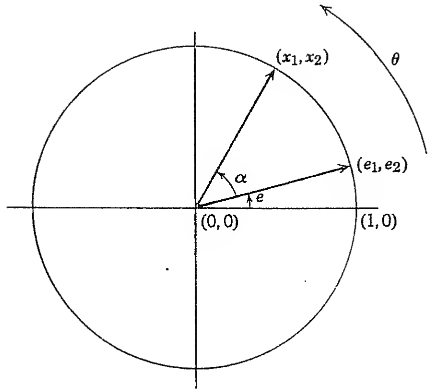


Figure 5 The error variable (e_1, e_2) describes the error angle e; the quantity θ is a rotation through an angle α.

Let the vector $(1, 0)$ be a *reference direction* on the plane R^2, and let a point $\mathbf{e} = (e_1, e_2)'$ on the unit circle give the *error angle e* measured positively from the reference direction (see Figure 5). Suppose the *error distribution* has been identified except perhaps for an *additional quantity* κ:

$$f(\mathbf{e}:\kappa)\,d\mathbf{e} = f(e_1, e_2 : \kappa)\,de.$$

The vector $\mathbf{e} = (e_1, e_2)'$ is restricted to the unit circle, and the differential $d\mathbf{e} = de$ measures length on the unit circle.

The physical quantity is the general direction of the property investigated. Let θ be this direction as designated by the rotation

$$\theta = \begin{bmatrix} \alpha_1 & -\alpha_2 \\ \alpha_2 & \alpha_1 \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

through an angle α from the reference direction. The quantity θ belongs to the *positive orthogonal group* of rotations on the plane:

$$G = \left\{ \begin{bmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{bmatrix} : \; 0 \le a < 2\pi \right\}.$$

For a single observation let $\mathbf{x} = (x_1, x_2)'$ be the measured direction. The model is

$$f(e_1, e_2 : \kappa)\,de,$$
$$\mathbf{x} = \theta\mathbf{e}.$$

For multiple observations let

$$X = (\mathbf{x}_1, \ldots, \mathbf{x}_n) = \begin{bmatrix} x_{11} \cdots x_{1n} \\ x_{21} \cdots x_{2n} \end{bmatrix}$$

designate n measured directions and let

$$E = (\mathbf{e}_1, \ldots, \mathbf{e}_n) = \begin{bmatrix} e_{11} \cdots e_{1n} \\ e_{21} \cdots e_{2n} \end{bmatrix}$$

designate the corresponding realized error or the corresponding error variable. This gives the

Measurement Model on the Circle

$$\prod f(\mathbf{e}_i : \kappa) \prod d\mathbf{e}_i,$$
$$X = \theta E.$$

The model has an error distribution describing the multiple measurement process, and it has a structural equation in which a realized error E has determined the relation between the measurement X and the quantity θ. The model is a conditional structural model with additional quantity κ.

Consider the effect of a transformation g:

$$gX = g(\mathbf{x}_1, \ldots, \mathbf{x}_n).$$

The transformation g takes the n points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ on the unit circle and rotates them through an angle a: The relative position of the points remains the same; the general placement on the circle is changed by a rotation through the angle a. To describe the position of the n points let

$$\mathbf{a}(X) = \begin{bmatrix} a_1(X) \\ a_2(X) \end{bmatrix} = \begin{bmatrix} \cos a(X) \\ \sin a(X) \end{bmatrix}$$

be the unit vector in the direction of the sum vector

$$\sum_1^n \mathbf{x}_i = \begin{bmatrix} \sum x_{1i} \\ \sum x_{2i} \end{bmatrix}.$$

Let $l(X)$ be the length of the sum vector. Then

$$l^2(X) = (\sum x_{1i})^2 + (\sum x_{2i})^2,$$

$$\mathbf{a}(X) = \frac{\sum_1^n \mathbf{x}_i}{l(X)} = \begin{bmatrix} \dfrac{\sum x_{1i}}{l(X)} \\ \dfrac{\sum x_{2i}}{l(X)} \end{bmatrix}.$$

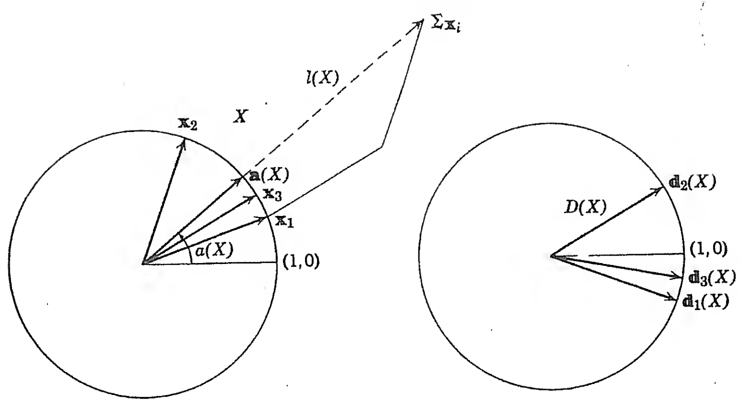


Figure 6 The array X; the location vector $\mathbf{a}(X)$; the reference array D.

See Figure 6. A transformation variable can be constructed as the rotation through the angle $a(X)$:

$$[X] = \begin{bmatrix} a_1(X) & -a_2(X) \\ a_2(X) & a_1(X) \end{bmatrix} = \begin{bmatrix} \cos a(X) & -\sin a(X) \\ \sin a(X) & \cos a(X) \end{bmatrix}.$$

The corresponding reference point is

$$D(X) = (\mathbf{d}_1(X), \ldots, \mathbf{d}_n(X)) = [X]^{-1}X = \begin{bmatrix} a_1(X) & a_2(X) \\ -a_2(X) & a_1(X) \end{bmatrix} X$$

$$= \begin{bmatrix} a_1(X)x_{11} + a_2(X)x_{21} & \cdots & a_1(X)x_{1n} + a_2(X)x_{2n} \\ -a_2(X)x_{11} + a_1(X)x_{21} & \cdots & -a_2(X)x_{1n} + a_1(X)x_{2n} \end{bmatrix}.$$

Note that the sum vector for $D(X)$ is in the reference direction and has length $l(D(X)) = l(X)$:

$$\sum \mathbf{d}_i(X) = [X]^{-1} \sum \mathbf{x}_i = [X]^{-1} l(X)\mathbf{a}(X) = l(X)\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The invariant differentials are the Euclidean differentials for the unit vectors involved (i.e., lengths on unit circles).

The conditional distribution of the error position $\mathbf{a} = \mathbf{a}(E)$ given the orbit D is

$$g_\kappa(\mathbf{a} : D)\, da = k_\kappa(D) \prod_1^n f(a_1 d_{1i} - a_2 d_{2i}, a_2 d_{1i} + a_1 d_{2i} : \kappa)\, da.$$

For an assumed value for the quantity κ the reduced structural model is

$$g_\kappa(\mathbf{a}:D(X))\,da,$$

$$a(X) = \alpha + a.$$

Note that $a(X)$ and a are the angles for the location directions $\mathbf{a}(X)$ and $\mathbf{a} = \mathbf{a}(E)$. See Figure 7. The structural distribution for the angle α conditional on the quantity κ is

$$k_\kappa(D) \prod_1^n f(\alpha_1 x_{1i} + \alpha_2 x_{2i}, -\alpha_2 x_{1i} + \alpha_1 x_{2i} : \kappa)\, d\alpha.$$

The marginal probability element for the orbit $D = D(E)$ with differentials at $D = D(X)$ is

$$\frac{1}{k_\kappa(D)} \cdot \frac{\prod dx_i}{da(X)},$$

and the marginal likelihood function for κ is

$$L(D:\kappa) = R^+(D)\,\frac{1}{k_\kappa(D)}.$$

A *normal error distribution for the circle* has been proposed:

$$f(e_1, e_2:\kappa)\, de = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa e_1\}\, de$$

$$= \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos e\}\, de.$$

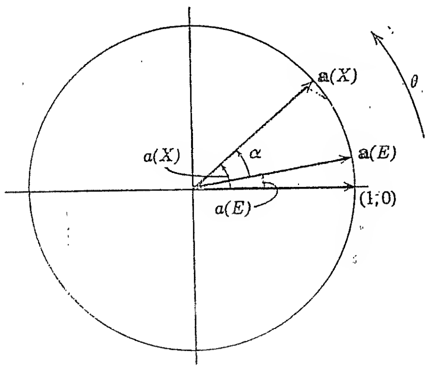


Figure 7 The observed position $\mathbf{a}(X)$; the unknown realized error position $\mathbf{a}(E)$; the unknown quantity θ.

The normalizing constant involves the imaginary Bessel function of zero order,

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \exp\{\kappa \cos e\}\, de.$$

The characteristic κ describes precision: With $\kappa = 0$ the distribution is uniform on the circle; with a large positive κ the distribution is concentrated on the circle near $(1, 0)$:

$$f(e_1, e_2:\kappa) = k' \exp\{\kappa(1 - \tfrac{1}{2}e^2 + \cdots)\}$$

$$\approx k'' \exp\{-\tfrac{1}{2}\kappa e^2\}.$$

The normal error distribution for the circle can be obtained from a symmetric normal distribution in the plane by conditioning to the unit circle (relative to the partition by circles about the origin).

The conditional distribution of the error position $\mathbf{a} = \mathbf{a}(E)$ given the orbit D is

$$g_\kappa(\mathbf{a}:D)\, da = k_\kappa(D)\,\frac{1}{(2\pi I_0(\kappa))^n} \exp\left\{\sum_1^n \kappa(a_1 d_{1i} - a_2 d_{2i})\right\} da$$

$$= k_\kappa(D)\,\frac{1}{(2\pi I_0(\kappa))^n} \exp\{l\kappa a_1\}\, da$$

$$= \frac{1}{2\pi I_0(l\kappa)} \exp\{l\kappa \cos a\}\, da.$$

The conditional distribution for the error position is also a normal error distribution for the circle but with *precision* $l\kappa = l(X)\kappa = l(D)\kappa$; the distribution depends on the orbital variable $D = D(X)$ *but only in terms of the real variable* $l = l(X)$. For an assumed value for the quantity κ the reduced structural model is

$$\frac{1}{2\pi I_0(l(X)\kappa)} \exp\{l(X)\kappa \cos a\}\, da,$$

$$a(X) = \alpha + a.$$

The structural distribution for the angle α conditional on the quantity κ is

$$\frac{1}{2\pi I_0(l(X)\kappa)} \exp\{l(X)\kappa \cos(a(X) - \alpha)\}\, d\alpha.$$

The marginal likelihood function is

$$L(D:\kappa) = R^+(D)\,\frac{1}{k_\kappa(D)} = R^+(D)\,\frac{2\pi I_0(l(D)\kappa)}{(2\pi I_0(\kappa))^n} = R^+(D)\,\frac{I_0(l(D)\kappa)}{I_0^n(\kappa)};$$

the marginal likelihood can be expressed as a ratio relative to $\kappa = 0$:

$$L^*(D:\kappa) = \frac{I_0(l(D)\kappa)}{I_0^n(\kappa)} \cdot \frac{I_0^n(0)}{I_0(0)} = \frac{I_0(l(D)\kappa)}{I_0^n(\kappa)}.$$

It is of interest that the marginal likelihood function depends on the orbit *only in terms of the real variable*

$$l = l(D).$$

It follows that the marginal distribution for the orbit D involves only l in its dependence on κ. And it follows as in the preceding section that, if

$$h(l:0)\, dl$$

is the marginal distribution of the length l with $\kappa = 0$, then

$$h(l:\kappa)\, dl = \frac{I_0(l\kappa)}{I_0^n(\kappa)} h(l:0)\, dl$$

is the marginal distribution of the length l for general κ.

The distribution of the length l of the sum vector based on the uniform distribution on the circle ($\kappa = 0$) is available from probability theory:

$$h(l:0)\, dl = l \int_0^\infty J_0(lu) J_0^n(u) u\, du \cdot dl, \qquad 0 < l < \infty,$$

where $J_0(u)$ is the Bessel function of zero order. The general distribution for l is then

$$h(l:\kappa)\, dl = \frac{I_0(l\kappa)}{I_0^n(\kappa)}\, l \int_0^\infty J_0(lu) J_0^n(u) u\, du \cdot dl, \qquad 0 < l < \infty.$$

*6 MARGINAL LIKELIHOOD: EXTENSIONS

The conditional structural model can be extended in two directions. In some contexts it may occur that the response variable X has been inappropriately expressed and that a transformation of X,

$$X_\lambda = l(X:\lambda),$$

dependent on some aspect of the additional quantity λ is in reality *the natural response variable*: A value X_λ of the natural response is produced by the transformation θ applied to a realized error value E. For each λ suppose that $X_\lambda = l(X:\lambda)$ is a one-to-one continuously differentiable function from the range of X onto the space $\mathfrak{X}$.

As a second extension suppose that the group G applies to the space $\mathfrak{X}$ in a way that depends on λ. Let θ as a transformation on $\mathfrak{X}$ be designated θ_λ, and let G as a group of transformations on $\mathfrak{X}$ be designated G_λ.

These two extensions give a generalized

Conditional Structural Model

$$f(E:\lambda)\, dE,$$
$$X_\lambda = \theta_\lambda E,$$

with additional quantity† λ. The model has an error variable E with distribution dependent on λ; and it has a structural equation in which a realized error value E is transformed by the quantity θ_λ in G_λ to give the natural response X_λ; the natural response† X_λ is related to the observed response X by the equation $X_\lambda = l(X:\lambda)$. If the additional quantity λ is known in value, then the conditional structural model is an ordinary structural model.

For analysis, let $G_\lambda X_\lambda$ be the orbit of X_λ under the transformation group G_λ,

$$G_\lambda X_\lambda = \{g_\lambda X_\lambda : \quad g_\lambda \in G_\lambda\}.$$

Let $[X_\lambda]_\lambda$ be a transformation variable relative to G_λ, and let $D_\lambda(X_\lambda)$ be the corresponding reference point:

$$X_\lambda = [X_\lambda]_\lambda D_\lambda(X_\lambda).$$

See Figure 8.

For an assumed value for the quantity λ the structural model produces a

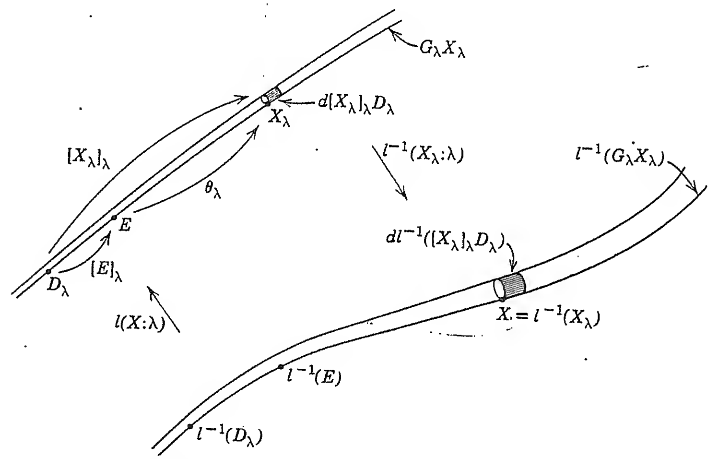


Figure 8 The orbit $G_\lambda X_\lambda$ passes through the natural response X_λ. The inverse image of the orbit $G_\lambda X_\lambda$ relative to the map $X_\lambda = l(X:\lambda)$ passes through the observed response X.

† The error distribution, the transformation group, and the form of the response depend on the quantity λ. The quantity λ may, indeed, have separate coordinates, one for each effect.

reduced structural model

$$g_\lambda([E]_\lambda : D_\lambda(X_\lambda))\, d[E]_\lambda,$$
$$[X_\lambda]_\lambda = \theta_\lambda[E]_\lambda,$$

where

$$g_\lambda([E]_\lambda : D_\lambda)\, d[E]_\lambda = k_\lambda(D_\lambda) f([E]_\lambda D_\lambda : \lambda) J_N([E]_\lambda : D_\lambda)\, d\mu([E]_\lambda).$$

The corresponding structural distribution for θ conditional on λ is

$$g_\lambda^*(\theta : X)\, d\theta = k_\lambda(D_\lambda) f(\theta_\lambda^{-1} X_\lambda : \lambda) J_N(\theta_\lambda^{-1}[X_\lambda]_\lambda : D_\lambda)\, \Delta([X_\lambda]_\lambda)\, d\nu(\theta).$$

For an assumed value for λ this distribution is the basis for inference concerning θ.

The probability element for E based on Euclidean volume is

$$f(E:\lambda)\, dE.$$

The conditional probability element for $[E]_\lambda$ given the orbit D_λ is

$$k_\lambda(D_\lambda) f([E]_\lambda D_\lambda : \lambda) \frac{J_N([E]_\lambda : D_\lambda)}{J_L([E]_\lambda)}\, d[E]_\lambda.$$

The marginal probability element for the orbit D_λ can then be obtained by dividing the full element by the conditional element

$$\frac{1}{k_\lambda(D_\lambda)} \frac{J_L([E]_\lambda)}{J_N([E]_\lambda : D_\lambda)} \frac{dE}{d[E]_\lambda}.$$

The marginal element at the point X_λ on the orbit D_λ rather than at the point E on the orbit D_λ is

$$\frac{1}{k_\lambda(D_\lambda)} \frac{J_L([X_\lambda]_\lambda)}{J_N([X_\lambda]_\lambda : D_\lambda)} \frac{dX_\lambda}{d[X_\lambda]_\lambda}.$$

The differential dX_λ can be expressed in terms of differential Euclidean volume for the observable variable:

$$dX_\lambda = \left| \frac{\partial l(X:\lambda)}{\partial X} \right| dX.$$

The differential $d[X_\lambda]_\lambda$ on the group can be expressed in terms of differential Euclidean volume in the L dimensions along the orbit:

$$d[X_\lambda]_\lambda = \frac{1}{K([X_\lambda]_\lambda : D_\lambda)}\, d[X_\lambda]_\lambda D_\lambda,$$

where

$$K([E]_\lambda : D_\lambda) = \left| \frac{\partial [E]_\lambda D_\lambda}{\partial [E]_\lambda} \right|.$$

The differential $d[X_\lambda]_\lambda D_\lambda$ along the orbit D_λ can then be expressed in terms of differential Euclidean volume for X along the inverse image of the orbit D_λ:

$$d[X_\lambda]_\lambda D_\lambda = \left| \frac{\partial l^{-1}([X_\lambda]_\lambda D_\lambda)}{\partial [X_\lambda]_\lambda D_\lambda} \right|^{-1} dl^{-1}([X_\lambda]_\lambda D_\lambda)$$

(Figure 8). Section 7 presents an example in which a combined effect of the preceding two differential adjustments can be calculated in a single step.

The marginal probability element for the orbit D_λ as based on cross-sectional Euclidean volume dv at the observed response value X is then

$$\frac{1}{k_\lambda(D_\lambda)} \cdot \frac{J_L([X_\lambda]_\lambda)\, K([X_\lambda]_\lambda : D_\lambda)}{J_N([X_\lambda]_\lambda : D_\lambda)} \cdot \frac{\left| \dfrac{\partial l(X:\lambda)}{\partial X} \right|}{\left| \dfrac{\partial l^{-1}([X_\lambda]_\lambda D_\lambda)}{\partial [X_\lambda]_\lambda D_\lambda} \right|^{-1}}\, dv.$$

The marginal likelihood for λ from the orbit D as observed at X is

$$\frac{R^+(D_\lambda)}{k_\lambda(D_\lambda)} \frac{J_L([X_\lambda]_\lambda)\, K([X_\lambda]_\lambda : D_\lambda)}{J_N([X_\lambda]_\lambda : D_\lambda)} \frac{\left| \dfrac{\partial l(X:\lambda)}{\partial X} \right|}{\left| \dfrac{\partial l^{-1}([X_\lambda]_\lambda D_\lambda)}{\partial [X_\lambda]_\lambda D_\lambda} \right|^{-1}}.$$

The marginal likelihood for λ is the basis for inference concerning the quantity λ.

7 THE TRANSFORMED REGRESSION MODEL

Consider again the regression model of Chapter Three. In some potential applications, familiarity with related systems may indicate a regression model with structural vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ and error distribution $f(e)\, de$, but the familiarity may leave doubt about the appropriate manner of expression for the response variable; for example, a response variable may be expressed, on first approach, as a variable y, yet detailed investigation may show that some transformed variable such as $\ln y$, $y^{1/2}$, or y^{-1} may be the appropriate variable for the regression model. A transformation having parameter λ that includes these three transformed variables is

$$\begin{aligned} y^{(\lambda)} &= y^\lambda, & \lambda &\neq 0, \\ &= \ln y & \lambda &= 0 \end{aligned}$$

(the transformation applies to a positive variable y).

Consider the response variable expressed as y, and suppose that familiarity with related systems indicates that the regression model can be applied to the transformed variable

$$y^{(\lambda)} = l(y, \lambda)$$

for some value of the *additional quantity* λ. Suppose that $l(y, \lambda)$ is a one-to-one continuously differentiable function and let

$$Y_\lambda = \begin{bmatrix} \mathbf{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_r' \\ \mathbf{y}^{(\lambda)'} \end{bmatrix} = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \\ y_1^{(\lambda)} & \cdots & y_n^{(\lambda)} \end{bmatrix}.$$

The *transformed regression model with additional quantity* λ can then be expressed as

$$f(E)\,dE,$$
$$Y_\lambda = \theta E.$$

Suppose that $n \geq r + 1$ and that $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent. For given λ the model is a structural model. The model then is a conditional structural model with additional quantity λ.

The conditional distribution of the error position $[E]$ given λ and the orbit $D(E) = D(Y_\lambda) = D_\lambda$ is

$$g([E]: D_\lambda)\, d[E] = k(D_\lambda) f([E] D_\lambda) \frac{s^n}{s^{r+1}}\, d[E]$$
$$= k(D_\lambda) \prod_1^n f\Big(\sum_u b_u v_{ui} + s d_i^{(\lambda)}\Big) s^{n-r-1} \prod db_u\, ds;$$

and for the case of normal error is

$$\frac{|VV'|^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}\mathbf{b}' VV' \mathbf{b}\}\, d\mathbf{b} \cdot \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} s^{n-r-1} \exp\left\{-\frac{s^2}{2}\right\} ds.$$

For an assumed value for λ the reduced structural model is

$$g([E]: D(Y_\lambda))\, d[E],$$
$$[Y_\lambda] = \theta[E].$$

For an assumed value for λ the structural distribution for θ is

$$k(D_\lambda) \prod_1^n f\left(\frac{y_i^{(\lambda)} - \sum \beta_u v_{ui}}{\sigma}\right) \left(\frac{s(\mathbf{y}^{(\lambda)})}{\sigma}\right)^n s^{-r}(\mathbf{y}^{(\lambda)}) \frac{\prod d\beta_u\, d\sigma}{\sigma};$$

and for the case of normal error is

$$\frac{|VV'|^{1/2}}{(2\pi\sigma^2)^{r/2}} \exp\{-\tfrac{1}{2}(\boldsymbol{\beta} - \mathbf{b}(\mathbf{y}^{(\lambda)}))' \frac{VV'}{\sigma^2} (\boldsymbol{\beta} - \mathbf{b}(\mathbf{y}^{(\lambda)}))\}\, d\boldsymbol{\beta}$$
$$\cdot \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} \left(\frac{s(\mathbf{y}^{(\lambda)})}{\sigma}\right)^{n-r-1} \exp\left\{-\frac{s^2(\mathbf{y}^{(\lambda)})}{2\sigma^2}\right\} \frac{s(\mathbf{y}^{(\lambda)})}{\sigma} \frac{d\sigma}{\sigma}.$$

For an assumed value for λ this provides the basis for inference concerning θ.

The marginal element for D_λ based on differentials at the point E is

$$\frac{f(E)\, dE}{g([E]: D_\lambda)\, d[E]} = \frac{1}{k(D_\lambda)} \frac{s^{r+1}(\mathbf{e})}{s^n(\mathbf{e})} \frac{dE}{d[E]},$$

and based on differentials at the point Y_λ is

$$\frac{1}{k(D_\lambda)} \frac{s^{r+1}(\mathbf{y}^{(\lambda)})}{s^n(\mathbf{y}^{(\lambda)})} \frac{dY_\lambda}{d[Y_\lambda]}.$$

The differential dY_λ can be expressed in terms of the differential dY:

$$(dy_1^{(\lambda)}, \ldots, dy_n^{(\lambda)}) = (dy_1, \ldots, dy_n) J(\mathbf{y}:\lambda),$$

where

$$J(\mathbf{y}:\lambda) = \begin{bmatrix} \dfrac{dy_1^{(\lambda)}}{dy_1} & & 0 \\ & \ddots & \\ 0 & & \dfrac{dy_n^{(\lambda)}}{dy_n} \end{bmatrix}$$

is the Jacobian matrix of the transformation, hence

$$dY_\lambda = |J(\mathbf{y}:\lambda)|\, dY.$$

The differential $d[Y_\lambda]$ can be expressed in terms of differential volume at Y along the inverse image of the orbit D_λ. For this the differential vector on the group can be related to the corresponding differential vector for the natural response,

$$(dy_1^{(\lambda)}, \ldots, dy_n^{(\lambda)}) = (db_1, \ldots, db_r, ds) D_\lambda;$$

and a differential vector for the natural response can be related to the corresponding differential vector at the observed response:

$$(dy_1, \ldots, dy_n) = (dy_1^{(\lambda)}, \ldots, dy_n^{(\lambda)}) J^{-1}(\mathbf{y}:\lambda).$$

The composite transformation is

$$(dy_1, \ldots, dy_n) = (db_1, \ldots, db_r, ds) D_\lambda J^{-1}(\mathbf{y}:\lambda).$$

Hence†

$$dl^{-1}([Y_\lambda]D_\lambda) = |D_\lambda J^{-2}(y:\lambda)D_\lambda'|^{1/2}\, d[Y_\lambda],$$

$$d[Y_\lambda] = |D_\lambda J^{-2}(y:\lambda)D_\lambda'|^{-1/2}\, dl^{-1}([Y_\lambda]D_\lambda).$$

The marginal element for D_λ can now be expressed in terms of cross-sectional Euclidean volume dv at the observed response Y:

$$\frac{1}{k(D_\lambda)}\,\frac{1}{s^{n-r-1}(y^{(\lambda)})}\,\frac{|J(y:\lambda)|}{|D_\lambda J^{-2}(y:\lambda)D_\lambda'|^{-1/2}}\,dv;$$

and for the case of normal error is

$$\frac{1}{A_{n-r}\,|D_\lambda D_\lambda'|^{1/2}}\,\frac{1}{s^{n-r-1}(y^{(\lambda)})}\,\frac{|J(y:\lambda)|}{|D_\lambda J^{-2}(y:\lambda)D_\lambda'|^{-1/2}}\,dv.$$

(note that $|VV'| = |D_\lambda D_\lambda'|$). The marginal likelihood function for D_λ is

$$\frac{R^+(D_\lambda)}{k(D_\lambda)s^{n-r-1}(y^{(\lambda)})}\,\frac{|J(y:\lambda)|}{|D_\lambda J^{-2}(y:\lambda)D_\lambda'|^{-1/2}};$$

and for normal error is

$$\frac{R^+(D_\lambda)}{s^{n-r-1}(y^{(\lambda)})}\,\frac{|J(y:\lambda)|}{|D_\lambda J^{-2}(y:\lambda)D_\lambda'|^{-1/2}}.$$

The marginal likelihood function is the basis for inference concerning the quantity λ.

NOTES AND REFERENCES

The likelihood function was promoted and developed in statistics by R. A. Fisher (1922, 1925, 1934, 1956). The concept received rather little attention from North American statisticans; notable exceptions are Birnbaum (1962) and some related papers. Another concept promoted by Fisher (1922, 1925, 1956), the concept of a sufficient statistic, received widespread attention, however. A close relationship exists between the two concepts; this was indicated informally by remarks in Fisher (1934) and was given explicit recognition by G. Barnard and J. W. Tukey at statistical meetings around 1960. The close relationship produces new and short methods of analysis for sufficiency in the classical model (Fraser, 1966a, c). The notation used for the likelihood function follows closely a notation proposed by J. Bondar.

† Let $y = h(u)$ be a continuously-differentiable transformation from R^L into R^N. At u let $J = (\partial h'/\partial u)$ be the $L \times N$ Jacobian matrix; then $dy' = du'J$. Let $J = TO$ be the triangular-orthogonal factorization of J (Section 6, Chapter Three); then $dy' = du'TO$. At y the differential changes are in the L-dimensional subspace spanned by the rows of O; these rows provide new axes for the subspace. In terms of the new axes the transformation is from du' to $du'T$. The Jacobian determinant is $|T| = |TT'|^{1/2} = |JJ'|^{1/2}$.

The concept of marginal likelihood was introduced in a preliminary form as residual likelihood (Fraser, 1964). For problems involving transformations of a response variable to obtain a linear model, Box and Cox (1964) applied likelihood methods and a modified Bayesian method—the Bayesian method using the theoretically desperate device of choosing a prior distribution on the basis of the observed response. The concept of marginal likelihood was introduced in an explicit form in Fraser (1967) and applied there to the regression model as in Section 7. The resulting marginal likelihood function avoids the approximations needed by Box and Cox and provides greater sensitivity to the data than can be obtained with the likelihood or modified Bayesian results. The use of the marginal likelihood methods in Section 3 avoids in a similar manner the need for the Bayesian methods in Box and Tiao (1962). A multivariate version of the Box and Cox problem has been analyzed by Fraser and L. M. Steinberg.

The distribution of the correlation coefficient was derived by Fisher (1915). The analysis of the composite response model in Section 4 produces, as a byproduct, a simple derivation of this distribution; it avoids the intermediate derivation of the covariance-matrix distribution. The method for deriving a general distribution by the *likelihood-modulation* of a special distribution was introduced by Watson (1956) and Watson and Williams (1956) in an analysis of distributions on a sphere (to be examined in Section 2, Chapter Five). A survey of expressions for the correlation-coefficient distribution is given by Hotelling (1953).

The normal distribution on the circle was proposed by Gumbel, Greenwood, and Durand (1953). The distribution had been used in other contexts and some related distribution theory solved (Kluyver, 1906; von Mises, 1918; Rayleigh, 1919). A survey of results concerning distributions on the circle is given by Stephens (1962).

Birnbaum, A. (1962), On the foundations of statistical inference, *J. Am. Statist. Assoc.*, **57**, 269–306.

Box, G. E. P. and Cox, D. R. (1964), An analysis of transformations, *J. Roy. Statist. Soc.*, **B26**, 211–243.

Box, G. E. P. and Tiao, G. C. (1962), A further look at robustness via Bayes's theorem, *Biometrika*, **49**, 419–432.

Fisher, R. A. (1915), Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika*, **10**, 507–521.

Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics, *Phil. Trans. Roy. Soc., London*. **A222**, 309–368; also as Paper 10, Fisher (1950).

Fisher, R. A. (1925), Theory of statistical estimation, *Proc. Cambridge Phil. Soc.*, **22**, 700–725; also as Paper 11, Fisher (1950).

Fisher, R. A. (1934), Two new properties of mathematical likelihood, *Proc. Roy. Soc.* (*London*), **A144**, 285–307; also as paper 24, Fisher (1950).
Fisher, R. A. (1950), *Contributions to Mathematical Statistics*, Wiley, New York.
Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, London.
Fraser, D. A. S. (1963), On the sufficiency and likelihood principles, *J. Am. Statist. Assoc.*, **58**, 641–647.
Fraser, D. A. S. (1964), On local inference and information, *J. Roy. Statist. Soc.*, **B26**, 253–260.
Fraser, D. A. S. (1966a), Sufficiency for regular models, *Sankhyā*, **A28**, 137–144.
Fraser, D. A. S. (1966b), On sufficiency and conditional sufficiency, *Sankhyā*, **A28**, 145–150.
Fraser, D. A. S. (1966c), Sufficiency for selection models, *Sankhyā*, **A28**, 329–334.
Fraser, D. A. S. (1967), Data transformations and the linear model, *Ann. Math. Statistics*, **38**, 1456–1465.
Gumbel, E. J., Greenwood, J. A., and Durand, D. (1953), The circular normal distribution: theory and tables, *J. Am. Statist. Assoc.*, **48**, 131–152.
Hotelling, H. (1953), New light on the correlation coefficient and its transforms, *J. Roy. Statist. Soc.*, **B15**, 193–218.
Kluyver, J. C. (1906), A local probability problem, *Proc. Acad. Amsterdam*, **8**, 341–350.
von Mises, R. (1918), Uber die Ganzzahligkeit der Atomgewichte und verwandte Fragen, *Physik. Z.*, **19**, 490–500.
Lord Rayleigh (1919), On the problem of random vibrations and random flights in one, two, and three dimensions, *Phil. Mag.*, (6) **37**, 321–347.
Stephens, M. A. (1962), *The statistics of directions, the Fisher and von Mises distributions*, Ph.D. thesis, University of Toronto.
Watson, G. S. (1956), Analysis of dispersion on a sphere, *Monthly Notices, Roy. Astron. Soc., Geophys. Suppl.*, **7**, 153–159.
Watson, G. S. and Williams, E. J. (1956), On the construction of significance tests on the circle and the sphere, *Biometrika*, **43**, 344–352.

PROBLEMS

1. Consider the *simple measurement model with additional shape quantity* β:

$$\prod_{i=1}^{n} f(e_i:\beta) \prod_{i=1}^{n} de_i,$$

$$\mathbf{x} = [\theta, 1]\mathbf{e}.$$

(i) Derive an expression for the structural distribution for θ given β.
(ii) Derive an expression for the marginal likelihood function for β.
(iii) For the case of normal error with scaling $\beta = \sigma$,

$$f(e:\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{e^2}{2\sigma^2}\right\},$$

obtain the marginal likelihood explicitly. Show that the value $\sigma = s_x$ for the additional quantity maximizes the marginal likelihood function. (L. M. Steinberg.)
(iv) (*Continuation*). Assume that s_x has the distribution of $\chi(n-1)^{-1/2}$ on $n-1$ degrees of freedom for $\sigma = 1$. Use the likelihood-modulation method at the end of Sections 4

and 5 to show that the general distribution of s_x is that of $\sigma\chi(n-1)^{-1/2}$ on $n-1$ degrees of freedom.

2. Consider the simple measurement model (Problem 1) with error distribution

$$f(e:\sigma)\,de = \frac{1}{\sigma}\exp\left\{\frac{e}{\sigma} - \exp\frac{e}{\sigma}\right\} de$$

and additional quantity σ (Problem 16 in Chapter One examined the measurement model based on a standardized form of the preceding error distribution.) (J. Whitney.)

(i) Derive the structural distribution for θ.
(ii) Derive the marginal likelihood function for σ. Determine the equation for obtaining the value of σ that maximizes the marginal likelihood function for σ.
(iii) Derive the classical model $f(x:\theta, \sigma)$ for the response variable x. Obtain equations for the (θ, σ) that maximizes the likelihood function based on the classical model.
(iv) Compare the equation for the appropriate σ value in (ii) and the equation for the inappropriate σ value in (iii).

3. Consider the multiplicative measurement model (Problem 19, Chapter One) with additional quantity δ:

$$\prod_{i=1}^{n} f(e_i:\delta) \prod_{i=1}^{n} de_i,$$

$$\mathbf{x} = [0, \theta]\mathbf{e}. \qquad \text{(H. Levenbach).}$$

(i) Derive an expression for the structural distribution for θ given δ.
(ii) Derive an expression for the marginal likelihood function for δ.
(iii) For the case of normal error with coefficient of variation δ

$$f(e:\delta) = \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(e-\delta)^2\}$$

obtain the marginal likelihood function explicitly. Find the value for δ that maximizes the marginal likelihood function. Use $s(\mathbf{x}) = (\sum x_i^2)^{1/2}$; let $t(\mathbf{x}) = \sqrt{n}\,\bar{x}/(\sum(x_i-\bar{x})^2)^{1/2} = \sqrt{n}\,\bar{d}/(1-n\bar{d}^2)^{1/2}$.
*(iv) Use the method of likelihood modulation (Sections 4, 5, and Notes and References) to obtain the general distribution of the essential variable $t(\mathbf{x})$ based on the orbit.

4. Consider a *composite measurement model with additional quantity* β:

$$f(E)\,dE = \prod_{1}^{n} f(e_{1i}, \ldots, e_{pi}:\beta) \prod_{1}^{n} (de_{1i} \cdots de_{pi}),$$

$$X = \theta E,$$

where

$$X = \begin{bmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \cdot \\ \cdot \\ \mathbf{x}_p' \end{bmatrix}, \qquad E = \begin{bmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \cdot \\ \cdot \\ \mathbf{e}_p' \end{bmatrix}, \qquad \theta = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \mu_1 & \sigma & & 0 \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \mu_p & 0 & & \sigma \end{bmatrix},$$

and θ is an element of the location-scale group

$$G = \left\{ \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_1 & c & & 0 \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ a_p & 0 & & c \end{bmatrix} : \begin{array}{c} -\infty < a_j < \infty \\ 0 < c < \infty \end{array} \right\}.$$

(i) Check that the transformations form a group.

(ii) In the pattern of Section 4 in Chapter Three and Problems 1, 2 in Chapter Three define a transformation variable and determine the reference point. Show that the model is a conditional structural model with additional quantity β.

(iii) Derive the invariant differentials and the modular function.

(iv) Derive the distribution for error position; derive the conditional structural distribution for θ given β.

(v) Derive the marginal likelihood function for β.

5 (*Continuation*). Consider the preceding composite measurement model and suppose the error distribution is standard normal:

$$f(e_1, \ldots, e_p) = \frac{1}{(2\pi)^{p/2}} \exp\left\{-\tfrac{1}{2} \sum e_j^2\right\}$$

(no additional quantity).

(i) Derive the distribution for error position $[E]$; use

$$s(X) = \left[\sum_j \sum_i (x_{ji} - \bar{x}_j)^2\right]^{1/2}$$

as a convenient scale variable.

(ii) Derive the structural distribution for $(\mu_1, \ldots, \mu_p, \sigma)$.

6. For the composite response model with normal error $(p = 2)$ in Section 4 derive the equation that must be solved to obtain the value of ρ that maximizes the marginal likelihood function.

7. For the normal distribution

$$\frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[(e_1 - \mu)^2 + e_2^2]\right\} de_1\, de_2$$

in the plane determine the conditional distribution given that $e_1^2 + e_2^2 = 1$; relate the κ of the normal distribution on the circle to the (μ, σ) of the preceding distribution in the plane (further grounds for the name "normal distribution on the circle").

8. Consider a *simple composite-measurement model* (known error scaling) *with additional quantity* β:

$$f(E)\, dE = \prod_1^n f(e_{1i}, \ldots, e_{pi} : \beta) \prod_1^n (de_{1i}, \ldots, de_{pi}),$$

$$X = \theta E,$$

where

$$X = \begin{bmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_p' \end{bmatrix}, \quad E = \begin{bmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e}_p' \end{bmatrix}, \quad \theta = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \mu_1 & 1 & & 0 \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ \mu_p & 0 & & 1 \end{bmatrix},$$

and θ is an element of the location group

$$G = \left\{ \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_1 & 1 & & 0 \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ a_p & 0 & & 1 \end{bmatrix} : -\infty < a_j < \infty \right\}.$$

(i) Check that the transformations form a group.

(ii) In the location pattern of Problem 4 define a transformation variable and determine the reference point. Show that the model is a conditional structural model with additional quantity β.

(iii) Derive the invariant differentials and the modular function.

(iv) Derive the distribution for the error position; derive the conditional structural distribution for θ given β.

(v) Derive the marginal likelihood function for β.

9. (*Continuation*). Consider the preceding measurement model and suppose the component error distribution is an uncorrelated normal:

$$f(e_1, \ldots, e_p) = \frac{1}{(2\pi)^{p/2}\sigma_1 \cdots \sigma_p} \exp\left\{-\tfrac{1}{2} \sum \frac{e_j^2}{\sigma_j^2}\right\}.$$

(i) Derive the distribution of the error position $[E]$. Derive the structural distribution for $(\mu_1, \ldots, \mu_p)$ given $(\sigma_1, \ldots, \sigma_p)$.

(ii) Derive the marginal likelihood function for $(\sigma_1, \ldots, \sigma_p)$.

10. Consider the simple measurement model with additional quantity λ,

$$\prod_{i=1}^n f(e_i) \prod_{i=1}^n de_i,$$

$$l(x_1 : \lambda) = [\theta, 1]e_1$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$l(x_n : \lambda) = [\theta, 1]e_n,$$

where $l(x : \lambda)$ is a continuously differentiable monotone function mapping the range of x

onto R^1 (x is the available response, $l(x:\lambda)$ is the natural response as based on the correct value for λ; see Section 6 or 7.)

(i) Derive an expression for the structural distribution for θ given λ.
(ii) Derive an expression for the marginal likelihood function for λ.
(iii) For the case of normal error,

$$f(e) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{e^2}{2\sigma^2}\right\},$$

obtain the marginal likelihood explicitly. Describe the value of λ that maximizes the marginal likelihood.

11. Consider the measurement model with additional quantity λ:

$$\prod_1^n f(e_i) \prod_1^n de_i,$$

$$l(x_1:\lambda) = [\mu, \sigma]e_1$$

$$\vdots$$

$$l(x_n:\lambda) = [\mu, \sigma]e_n,$$

where $l(x:\lambda)$ is a continuously differentiable monotone function mapping the range of x onto R^1 (see Section 6 or 7).

(i) Derive an expression for the structural distribution for $[\mu, \sigma]$ given λ.
(ii) Derive an expression for the marginal likelihood function for λ.
(iii) For the case of normal error,

$$f(e) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{e^2}{2}\right\},$$

obtain the marginal likelihood explicitly. Describe the value of λ that maximizes the marginal likelihood.

12. Consider the simple regression model with additional quantity λ:

$$f(E)\,dE = \prod_1^n f(e_i) \prod_1^n de_i,$$

$$Y_\lambda = \theta E,$$

where

$$Y_\lambda = \begin{bmatrix} v_1' \\ \cdot \\ \cdot \\ v_r' \\ y^{(\lambda)\prime} \end{bmatrix}, \quad E = \begin{bmatrix} v_1' \\ \cdot \\ \cdot \\ v_r' \\ e' \end{bmatrix}, \quad \theta = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ 0 & & & 1 & 0 \\ \beta_1 & \cdots & & \beta_r & 1 \end{bmatrix},$$

$$y_i^{(\lambda)} = l(y_i:\lambda),$$

and $y^{(\lambda)} = l(y:\lambda)$ is a one-to-one continuously differentiable map carrying y to $y^{(\lambda)}$ (cf. Problem 17 in Chapter Three).

(i) Derive an expression for the structural distribution for θ given λ.
(ii) Derive an expression for the marginal likelihood function for λ.

13 (*Continuation*). Consider the simple regression model with normal component error

$$f(e)\,de = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left\{-\frac{e^2}{2\sigma_0^2}\right\} de$$

(cf. Problem 18, Chapter Three).

(i) Derive the structural distribution for θ.
(ii) Derive the marginal likelihood function for λ.

***14.** Consider the simple regression model

$$\frac{1}{(2\pi)^{n/2}\sigma_0^n} \exp\left\{-\frac{1}{2}\sum \frac{e_i^2}{\sigma_0^2}\right\} \prod de_i,$$

$$Y = \theta_\lambda E,$$

where

$$Y = \begin{bmatrix} v_1'(\lambda) \\ \cdot \\ \cdot \\ v_r'(\lambda) \\ y' \end{bmatrix}, \quad E = \begin{bmatrix} v_1'(\lambda) \\ \cdot \\ \cdot \\ v_r'(\lambda) \\ e' \end{bmatrix}, \quad \theta_\lambda = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ 0 & & & 1 & 0 \\ \beta_1 & \cdots & & \beta_r & 1 \end{bmatrix}$$

and λ is an additional quantity; note that θ_λ relocates relative to $V(\lambda)$.

(i) Determine the structural distribution for $\boldsymbol{\beta}$ given λ.
(ii) Determine the marginal likelihood for λ; describe the λ value that maximizes the marginal likelihood.

***15.** Consider the regression model

$$\frac{1}{(2\pi)^{n/2}} \exp\left\{-\tfrac{1}{2}\sum e_i^2\right\} \prod de_i,$$

$$Y = \theta_\lambda E,$$

where

$$Y = \begin{bmatrix} v_1'(\lambda) \\ \cdot \\ \cdot \\ v_r'(\lambda) \\ y' \end{bmatrix}, \quad E = \begin{bmatrix} v_1'(\lambda) \\ \cdot \\ \cdot \\ v_r'(\lambda) \\ e' \end{bmatrix}, \quad \theta_\lambda = \begin{bmatrix} 1 & & & 0 & 0 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ 0 & & & 1 & 0 \\ \beta_1 & \cdots & & \beta_r & \sigma \end{bmatrix},$$

and λ is an additional quantity; note that θ_λ relocates relative to $V(\lambda)$.

(i) Determine the structural distribution for $\boldsymbol{\beta}$, σ given λ.
(ii) Determine the marginal likelihood for λ; describe the λ value that maximizes the marginal likelihood.

CHAPTER FIVE

Marginal Analysis

In some applications a structural model

$$f(E)\,dE,$$

$$X = \theta E$$

may have an error distribution that has symmetries with respect to its transformation group G. Let H be the set of transformations that do not alter the error distribution:

$$H = \{g\colon\ f(g^{-1}E)\,dg^{-1}E = f(E)\,dE,\quad g \in G\}$$
$$= \{g\colon\ f(g^{-1}E) = f(E),\quad g \in G\};$$

The set H consists of the transformations g for which the variable gE has the same distribution as the variable E. Clearly, if g_1 and g_2 are in H, then g_2g_1 and g_1^{-1} are in H. It follows that H is a subgroup of G, the *stabilizer subgroup* for the error distribution $f(E)\,dE$.

Consider a transformation τ in G and a transformation φ in the stabilizer subgroup H. The composite transformation $\tau\varphi$ applied to a value E from the error variable would give the same response value $\tau\varphi E$ as the transformation τ applied to the value φE of an equivalent error variable. It follows that the transformations $\tau\varphi$ for various φ in H are *equivalent* transformations in the group G. These equivalent transformations form a *left coset* τH *of the stabilizer subgroup* H (see Figure 1).

In a typical application the left cosets τH of a subgroup H can be indexed by the elements of a complementing subgroup H_2 in G. This was mentioned briefly at the beginning of Section 7 in Chapter Two.

Now consider inference for a structural model with a stabilizer subgroup H. The general inference concerning the quantity θ can be expressed in terms of the structural distribution:

$$g^*(\theta : X)\,d\theta = k([X]^{-1}X)\,f(\theta^{-1}X)\,J_N(\theta^{-1}X)\,\Delta([X])\,d\nu(\theta).$$

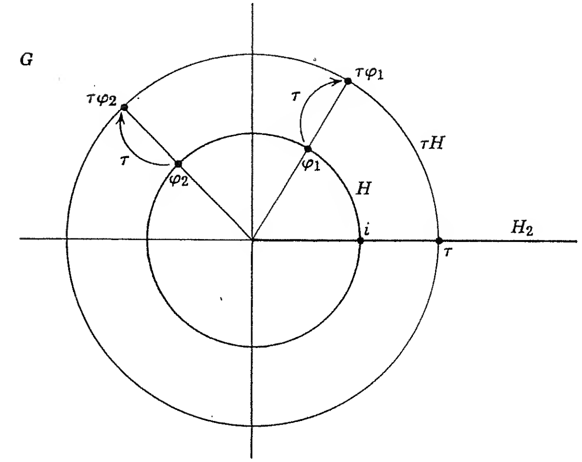


Figure 1 The stabilizer subgroup H for the error distribution; the left coset τH; a complementing subgroup H_2.

All values for θ in a left coset τH, however, are equivalent. It is natural then to present inference in terms of the essential quantity τ in the complementing subgroup H_2—to obtain the *marginal structural distribution* for the essential quantity τ. The general formulation for marginal structural distributions was given in Section 8, Chapter Two. In this chapter some structural models with stabilizer subgroups are examined. The marginal structural distributions for the essential quantities are derived directly.

1 A COMPOSITE MEASUREMENT MODEL WITH KNOWN SCALING

Consider a measurement process on the Euclidean plane R^2. Let (e_1, e_2) be the error variable with known distribution

$$f(e_1, e_2)\,de_1\,de_2$$

on the Euclidean plane. Let (x_1, x_2) designate a measurement; let (μ_1, μ_2) designate the quantity being measured and φ designate an unknown angle, the angle through which an error value (e_1, e_2) is rotated to give the difference between the measurement and the quantity.

The model for a single measurement can be expressed as

$$f(e_1, e_2)\,de_1\,de_2,$$

$$\begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \mu_1 & \cos(\varphi) & -\sin(\varphi) \\ \mu_2 & \sin(\varphi) & \cos(\varphi) \end{bmatrix} \begin{bmatrix} 1 \\ e_1 \\ e_2 \end{bmatrix}.$$

The transformation

$$\theta = \begin{bmatrix} 1 & 0 & 0 \\ \mu_1 & \cos(\varphi) & -\sin(\varphi) \\ \mu_2 & \sin(\varphi) & \cos(\varphi) \end{bmatrix}$$

is an element of the *translation-rotation group*:

$$G = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ a_1 & \cos(h) & -\sin(h) \\ a_2 & \sin(h) & \cos(h) \end{bmatrix} : \begin{array}{c} -\infty < a_j < \infty \\ 0 \le h < 2\pi \end{array} \right\}.$$

The model for n measurements is

$$\prod f(e_{1i}, e_{2i}) \prod de_{1i}\, de_{2i},$$
$$X = \theta E,$$

where

$$X = \begin{bmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \end{bmatrix}, \qquad E = \begin{bmatrix} 1 & \cdots & 1 \\ e_{11} & \cdots & e_{1n} \\ e_{21} & \cdots & e_{2n} \end{bmatrix}.$$

The matrix X can be viewed as a point in R^{2n} or as n numbered points in R^2. Toward defining a transformation variable, consider X as n numbered points in R^2: if the points for X can be carried point-for-point into the points for $\tilde{X}$ by translation and rotation, then X and $\tilde{X}$ are on the same orbit. As a transformation variable consider

$$[X] = \begin{bmatrix} 1 & 0 & 0 \\ \bar{x}_1 & \cos(h(X)) & -\sin(h(X)) \\ \bar{x}_2 & \sin(h(X)) & \cos(h(X)) \end{bmatrix},$$

where

$$(\cos(h(X)), \sin(h(X))) = \left(\frac{x_{11} - \bar{x}_1}{r_1(X)}, \frac{x_{21} - \bar{x}_2}{r_1(X)}\right),$$
$$r_1^2(X) = (x_{11} - \bar{x}_1)^2 + (x_{21} - \bar{x}_2)^2.$$

Applying the transformation $[X]^{-1}$ to the point X gives

$$[X]^{-1}X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_1(X) & d_{12} & \cdots & d_{1n} \\ 0 & d_{22} & \cdots & d_{2n} \end{bmatrix} = D(X),$$

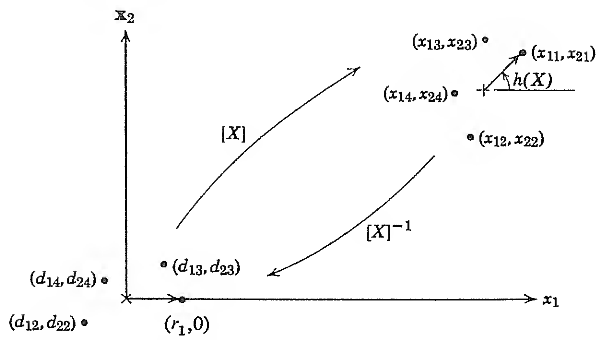


Figure 2 The general *point* X; the reference point $D(X)$.

where

$$\begin{bmatrix} d_{1i} \\ d_{2i} \end{bmatrix} = \begin{bmatrix} \cos(h(X))(x_{1i} - \bar{x}_1) + \sin(h(X))(x_{2i} - \bar{x}_2) \\ -\sin(h(X))(x_{1i} - \bar{x}_1) + \cos(h(X))(x_{2i} - \bar{x}_2) \end{bmatrix},$$

The transformation $[X]^{-1}$ translates the n points so that the center of gravity is at the origin and then rotates the points so that the vector from the center of gravity to the point (x_{11}, x_{12}) is in the direction of the first axis (see Figure 2). The model is a structural model for $n \ge 2$ (delete points X having $x_{11} = \bar{x}_1, x_{21} = \bar{x}_2$).

Translation and rotation do not change Euclidean volume. Accordingly, the invariant differential on R^{2n} is the Euclidean differential itself, and similarly the invariant differential on the group is the Euclidean differential $da_1\, da_2\, dh$.

The distribution of the error position $[E]$ given the orbit $D(E) = D(X) = D$ is

$$k(D) \prod_1^n f(\bar{e}_1 + \cos(h)\, d_{1i} - \sin(h)\, d_{2i},$$
$$\bar{e}_2 + \sin(h)\, d_{1i} + \cos(h)\, d_{2i})\, d\bar{e}_1\, d\bar{e}_2\, dh.$$

The structural distribution for θ is

$$k(D) \prod_1^n f(\cos(\varphi)(x_{1i} - \mu_1) + \sin(\varphi)(x_{2i} - \mu_2),$$
$$-\sin(\varphi)(x_{1i} - \mu_1) + \cos(\varphi)(x_{2i} - \mu_2))\, d\mu_1\, d\mu_2\, d\varphi.$$

Suppose now that the error distribution is rotationally symmetric about the origin:

$$f(\cos(h)e_1 + \sin(h)e_2, -\sin(h)e_1 + \cos(h)\, e_2) = f(e_1, e_2)$$

for $0 \le h < 2\pi$. The stabilizer subgroup for the error distribution is the rotation group

$$H = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(h) & -\sin(h) \\ 0 & \sin(h) & \cos(h) \end{pmatrix} : \quad 0 \le h < 2\pi \right\}.$$

A complementing subgroup in G is the translation group

$$H_2 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ a_1 & 1 & 0 \\ a_2 & 0 & 1 \end{pmatrix} : \quad -\infty < a_j < \infty \right\}.$$

The marginal structural distribution for the location quantity (μ_1, μ_2) is obtained by integration:

$$k(D)\int_0^{2\pi} \prod_1^n f\big(\cos(\varphi)(x_{1i} - \mu_1) + \sin(\varphi)(x_{2i} - \mu_2),$$
$$-\sin(\varphi)(x_{1i} - \mu_1) + \cos(\varphi)(x_{2i} - \mu_2)\big)\, d\varphi\, d\mu_1\, d\mu_2$$
$$= 2\pi k(D) \prod_1^n f(x_{1i} - \mu_1, x_{2i} - \mu_2)\, d\mu_1\, d\mu_2.$$

2 THE MEASUREMENT MODEL ON A SPHERE

Consider a geologist measuring the direction of crystallization in samples of a certain rock structure, or a geophysicist measuring the direction of a magnetic field, or an astronomer measuring the direction of incoming radio waves. The quantity being measured is a direction. A direction can be recorded as a unit vector in R^3, as a point on the unit sphere. The *measurement model on the sphere* can be used to analyze the data.

Let $\mathbf{e} = (e_1, e_2, e_3)'$ be the error variable, a point on the unit sphere in R^3. And suppose that $\mathbf{e}$ records *error* in relation to a *reference direction* $(1, 0, 0)$ in R^3 (see Figure 3). Also suppose that the *error distribution* has been identified except perhaps for an additional quantity κ:

$$f(e_1, e_2, e_3 : \kappa)\, d\mathbf{e};$$

the vector $\mathbf{e}$ is restricted to the unit sphere, and the differential $d\mathbf{e}$ measures area on the unit sphere.

The physical quantity is the general direction of the property being investigated *and* the angle of rotation applied to an error value. Let θ be the quantity describing the general direction and the error rotation:

$$\theta = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{pmatrix} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3).$$

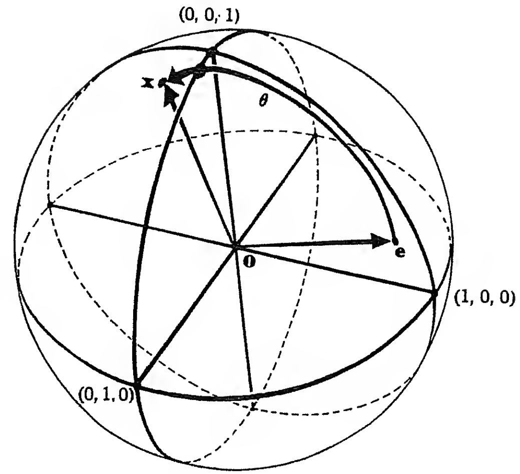


Figure 3 The error variable $\mathbf{e} = (e_1, e_2, e_3)'$. The measurement $\mathbf{x} = (x_1\ x_2\ x_3)$. The quantity θ, a rotation on the sphere.

The quantity θ is an element of the *rotation group* on R^3:

$$G = \left\{ O = \begin{pmatrix} o_{11} & o_{12} & o_{13} \\ o_{21} & o_{22} & o_{23} \\ o_{31} & o_{32} & o_{33} \end{pmatrix} : \quad \begin{matrix} O'O = I \\ |O| = 1 \end{matrix} \right\}.$$
$$= (\mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3)$$

For a single measurement let $\mathbf{x} = (x_1, x_2, x_3)'$ be the measured direction. The model is

$$f(e_1, e_2, e_3 : \kappa)\, d\mathbf{e},$$
$$\mathbf{x} = \theta\mathbf{e}.$$

For multiple measurements let X designate n measured directions,

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \\ x_{31} & \cdots & x_{3n} \end{pmatrix},$$

and let E designate the corresponding realized error or the corresponding error variable,

$$E = \begin{pmatrix} e_{11} & \cdots & e_{1n} \\ e_{21} & \cdots & e_{2n} \\ e_{31} & \cdots & e_{3n} \end{pmatrix}.$$

The model is

Measurement Model on the Sphere

$$f(E:\kappa)\,dE = \prod_1^n f(e_{1i}, e_{2i}, e_{3i}:\kappa) \prod_1^n d\mathbf{e}_i,$$

The model has an error distribution describing the multiple measurement process, and it has a structural equation in which a realized error E has determined the relation between the measurement X and the quantity θ. For $n \geq 2$ the model is a conditional structural model with additional quantity κ.

Consider the effect of a transformation g:

$$gX = g(\mathbf{x}_1, \ldots, \mathbf{x}_n).$$

The transformation g rotates the n points on the surface of the sphere but maintains their *relative* positions. Toward defining a transformation variable let $\mathbf{o}_1(X)$ be a unit vector in the direction of the sum vector $\Sigma\, \mathbf{x}_i$:

$$\mathbf{o}_1(X) = \begin{pmatrix} o_{11}(X) \\ o_{21}(X) \\ o_{31}(X) \end{pmatrix} = \begin{pmatrix} \dfrac{\sum x_{1i}}{l(X)} \\ \dfrac{\sum x_{2i}}{l(X)} \\ \dfrac{\sum x_{3i}}{l(X)} \end{pmatrix},$$

where

$$l^2(X) = (\textstyle\sum x_{1i})^2 + (\sum x_{2i})^2 + (\sum x_{3i})^2.$$

Let $\mathbf{o}_2(X)$ be the unit residual vector for $\mathbf{x}_1$ after regression on $\Sigma_1^n\, \mathbf{x}_i$. And let $\mathbf{o}_3(X)$ be the unique unit vector that then completes a right triad of three orthonormal vectors. As a transformation variable consider

$$[X] = (\mathbf{o}_1(X), \mathbf{o}_2(X), \mathbf{o}_3(X)) = O(X).$$

The application of the transformation $[X]^{-1}$ to the point X gives

$$[X]^{-1}X = (\mathbf{d}_1(X), \ldots, \mathbf{d}_n(X))$$
$$= \begin{pmatrix} d_{11}(X) & d_{12}(X) & \cdots & d_{1n}(X) \\ d_{21}(X) & d_{22}(X) & \cdots & d_{2n}(X) \\ 0 & d_{32}(X) & \cdots & d_{3n}(X) \end{pmatrix} = D(X);$$

the sum vector for $D(X)$ is

$$\sum_1^n \mathbf{d}_i(X) = \sum_1^n [X]^{-1}\mathbf{x}_i = [X]^{-1} l(X)\mathbf{o}_1(X)$$
$$= l(X)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

The transformation $[X]^{-1}$ carries the original sum vector into a vector along the reference vector $(1, 0, 0)$ and carries the first vector $\mathbf{x}_1$ into the plane of the first two axes. The point $D(X)$ is a reference point, and the transformation $[X]$ gives the position of X relative to the reference point.

The Euclidean area elements on the surface of the sphere are invariant under the rotations. For a Euclidean volume element dO on the group let $d\mathbf{o}_1$ be Euclidean area on the unit sphere in R^3 for $\mathbf{o}_1$; and let $d\mathbf{o}_2$ be Euclidean length on the unit circle in R^2 for $\mathbf{o}_2$ (a unit vector orthogonal to $\mathbf{o}_1$). The differential $d\mathbf{o}_1\, d\mathbf{o}_2$ is invariant under left and right matrix multiplication by positive orthogonal matrices.

The conditional distribution of the error position $[E] = O(E)$ given the orbit D is

$$g_\kappa(O:D)\,dO = k_\kappa(D) \prod_1^n f(O\mathbf{d}_i:\kappa)\, d\mathbf{o}_1\, d\mathbf{o}_2.$$

For an assumed value for the quantity κ the reduced structural model is

$$g_\kappa(O:D)\,dO,$$
$$O(X) = \theta O.$$

The structural distribution for the rotation θ conditional on the quantity κ is

$$k_\kappa(D) \prod_1^n f(\boldsymbol{\omega}_1'\mathbf{x}_i, \boldsymbol{\omega}_2'\mathbf{x}_i, \boldsymbol{\omega}_3'\mathbf{x}_i:\kappa)\, d\boldsymbol{\omega}_1\, d\boldsymbol{\omega}_2,$$

where for example $\boldsymbol{\omega}_1'\mathbf{x}_i$ is $(\boldsymbol{\omega}_1, \mathbf{x}_i)$, the inner product of $\boldsymbol{\omega}_1$ and $\mathbf{x}_i$.

The marginal probability element for D is

$$\frac{1}{k_\kappa(D)} \frac{\prod_1^n d\mathbf{x}_i}{d\mathbf{o}_1(X)\, d\mathbf{o}_2(X)};$$

and the marginal likelihood for κ is

$$R^+(D)\frac{1}{k_\kappa(D)}.$$

3 THE MEASUREMENT MODEL ON THE SPHERE: NORMAL ERROR

A *normal error distribution for the sphere* has been proposed:

$$f(e_1, e_2, e_3 : \kappa)\, de = \frac{\kappa}{4\pi \sinh(\kappa)} \exp\{\kappa e_1\}\, de.$$

The normalizing constant is easily checked:

$$\begin{aligned}\int \exp\{\kappa e_1\}\, de &= \int_0^{\pi} \exp\{\kappa \cos(e)\} 2\pi \sin(e)\, de \\ &= \int_{-1}^{+1} \exp\{\kappa t\} 2\pi\, dt \\ &= \frac{2\pi}{\kappa}(\exp\{\kappa\} - \exp\{-\kappa\}) = \frac{4\pi \sinh(\kappa)}{\kappa};\end{aligned}$$

the variable e is used to designate the angle between $\mathbf{e}$ and $(1, 0, 0)$. The quantity κ describes precision: with $\kappa = 0$, the distribution is uniform on the sphere; with κ large the distribution is concentrated near $(1, 0, 0)$ (compare with the normal example in Section 5, Chapter Four).

The conditional distribution of the error position $O(E)$, given the orbit D, is

$$\begin{aligned}g_\kappa(O : D)\, dO &= k_\kappa(D) \frac{\kappa^n}{(4\pi \sinh(\kappa))^n} \exp\left\{\sum_1^n \kappa(o_{11}, o_{12}, o_{13})\mathbf{d}_i\right\} d\mathbf{o}_1\, d\mathbf{o}_2 \\ &= \frac{\kappa l(X)}{4\pi \sinh(\kappa l(X))} \exp\{\kappa l(X) o_{11}\}\, d\mathbf{o}_1 \cdot \frac{d\mathbf{o}_2}{2\pi}.\end{aligned}$$

The conditional error distribution has two components: The distribution of $\mathbf{o}_1$ is a normal distribution on the sphere with precision $\kappa l(X)$; the distribution of $\mathbf{o}_2$ is uniform on the unit circle orthogonal to $\mathbf{o}_1$.

The structural distribution for the rotation θ conditional on the quantity κ is

$$\begin{aligned}&\frac{\kappa l(X)}{4\pi \sinh(\kappa l(X))} \exp\{\kappa l(X)(\omega_{11} o_{11}(X) + \omega_{21} o_{21}(X) + \omega_{31} o_{31}(X))\}\, d\boldsymbol{\omega}_1 \cdot \frac{d\boldsymbol{\omega}_2}{2\pi} \\ &\quad = \frac{\kappa l(X)}{4\pi \sinh(\kappa l(X))} \exp\{\kappa l(X) \boldsymbol{\omega}_1' \mathbf{o}_1(X)\}\, d\boldsymbol{\omega}_1 \cdot \frac{d\boldsymbol{\omega}_2}{2\pi}.\end{aligned}$$

The structural distribution for $\boldsymbol{\omega}_1$ is the normal distribution with precision $\kappa l(X)$ as relocated in the direction $\mathbf{o}_1(X)$ of the sum vector $\Sigma_1^n \mathbf{x}_i$; and the structural distribution for $\boldsymbol{\omega}_2$ is uniform on the unit circle orthogonal to $\boldsymbol{\omega}_1$.

The marginal likelihood function for κ is

$$R^+(D) \frac{4\pi \sinh(\kappa l(X)) 2\pi}{\kappa l(X)} \cdot \frac{\kappa^n}{(4\pi \sinh(\kappa))^n} = R^+(D) \frac{\sinh(\kappa l(X))}{\sinh^n(\kappa)} \kappa^{n-1}.$$

The marginal likelihood function depends on the orbit $D = D(E) = D(X)$ but *only in terms of* the length $l = l(E) = l(X)$. Correspondingly, the distribution for the orbit $D = D(E)$ as it depends on κ involves only the length $l = l(E)$. It follows then that the general distribution for l,

$$h(l : \kappa)\, dl,$$

can be obtained from a special distribution such as

$$h(l : 0)\, dl$$

by likelihood-modulation:

$$h(l : \kappa)\, dl = \frac{(\sinh(\kappa l)/\sinh^n(\kappa))\kappa^{n-1}}{l} h(l : 0)\, dl.$$

(Compare with Section 4 in Chapter Four.) The distribution of $l(E)$ for a uniform distribution on the sphere is available from probability theory:

$$h(l : 0)\, dl = \frac{2l}{\pi} \int_0^\infty \frac{\sin^n(t) \sin(lt)}{t^{n-1}}\, dt \cdot dl = \frac{l}{2^{n-1}} \varphi_n(l)\, dl,$$

as the distribution of the length of the sum of r random unit vectors in R^3; the function $\varphi_n(l)$ is

$$\varphi_n(l) = \frac{1}{(n-2)!} \sum_{s=0}^n \binom{n}{s} (-1)^s (n - l - 2s)_+^{n-2},$$

where

$$\begin{aligned}(t)_+ &= t, \quad &\text{if } t \geq 0, \\ &= 0, \quad &\text{if } t < 0.\end{aligned}$$

The general distribution of l is then

$$\begin{aligned}h(l : \kappa)\, dl &= \frac{\sinh(\kappa l)}{\sinh^n(\kappa)} \kappa^{n-1} \frac{2}{\pi} \int_0^\infty \frac{\sin^n(t) \sin(lt)}{t^{n-1}}\, dt \cdot dl \\ &= \frac{\sinh(\kappa l)}{\sinh^n(\kappa)} \kappa^{n-1} \frac{1}{2^{n-1}} \varphi_n(l)\, dl.\end{aligned}$$

4 THE MULTIVARIATE MODEL†

Consider a system with p response variables $y_1, \ldots, y_p$. Suppose the internal error, as it affects the responses, has been identified and can be described by

† The analysis of the multivariate model (Sections 4, 5, 6, 7) depends on definitions and notation in Sections 10 and 11 of Chapter Three. The multivariate model may be omitted on a first reading; the model, however, has a central place in mathematical statistics.

p error variables $e_1, \ldots, e_p$ with a known distribution on R^p. Let $\mu_1, \ldots, \mu_p$ be the general levels for the p response variables. And suppose that the error variables affect the response levels by linear distortion: for the jth response let $\gamma_{jj'}$ be the coefficient applied to the j'th error. A realized error vector and the corresponding response vector are then connected by the equations:

$$\begin{aligned} y_1 &= \mu_1 + \gamma_{11}e_1 + \cdots + \gamma_{1p}e_p \\ &\;\vdots \\ y_p &= \mu_p + \gamma_{p1}e_1 + \cdots + \gamma_{pp}e_p, \end{aligned}$$

or by the matrix equation

$$\begin{pmatrix} 1 \\ y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \mu_1 & \gamma_{11} & \cdots & \gamma_{1p} \\ \vdots & \vdots & & \vdots \\ \mu_p & \gamma_{p1} & \cdots & \gamma_{pp} \end{pmatrix} \begin{pmatrix} 1 \\ e_1 \\ \vdots \\ e_p \end{pmatrix}.$$

Now consider n performances of the system and let $\mathbf{y}_1' = (y_{11}, \ldots, y_{1n})$ be the observations for the first response, ..., and $\mathbf{y}_p' = (y_{p1}, \ldots, y_{pn})$ be the observations for the pth response. The system and the n performances can then be described by the

Affine Multivariate Model

$$\prod_1^n f(e_{1i}, \ldots, e_{pi}) \prod_1^n de_{1i} \cdots de_{pi},$$

$$\begin{pmatrix} \mathbf{1}' \\ \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_p' \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \mu_1 & \gamma_{11} & \cdots & \gamma_{1p} \\ \vdots & \vdots & & \vdots \\ \mu_p & \gamma_{p1} & \cdots & \gamma_{pp} \end{pmatrix} \begin{pmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix}$$

The model has an error distribution with $\mathbf{e}_1', \ldots, \mathbf{e}_p'$ as variables; and it has a structural equation in which realized errors $\mathbf{e}_1', \ldots, \mathbf{e}_p'$ have determined the relation between the observations and the quantities.

For matrix notation let

$$Y = \begin{pmatrix} \mathbf{1}' \\ \hline \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_p' \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ \hline y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{p1} & \cdots & y_{pn} \end{pmatrix} = \begin{pmatrix} \mathbf{1}' \\ \hline \underline{Y} \end{pmatrix},$$

$$E = \begin{pmatrix} \mathbf{1}' \\ \hline \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ \hline e_{11} & \cdots & e_{1n} \\ \vdots & & \vdots \\ e_{p1} & \cdots & e_{pn} \end{pmatrix} = \begin{pmatrix} \mathbf{1}' \\ \hline \underline{E} \end{pmatrix},$$

$$\theta = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline \mu_1 & \gamma_{11} & \cdots & \gamma_{1p} \\ \vdots & \vdots & & \vdots \\ \mu_p & \gamma_{p1} & \cdots & \gamma_{pp} \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline \boldsymbol{\mu} & \Gamma \end{array}\right).$$

The affine multivariate model can now be written:

$$f(E)\,dE,$$

$$Y = \theta E.$$

The transformation θ is an element of the *positive affine group* on R^p:

$$G = \left\{ g = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline a_1 & c_{11} & \cdots & c_{1p} \\ \vdots & \vdots & & \vdots \\ a_p & c_{p1} & \cdots & c_{pp} \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline \mathbf{a} & C \end{array}\right) : \begin{array}{l} -\infty < a_j < \infty \\ -\infty < c_{jj'} < \infty \\ |C| > 0 \end{array} \right\},$$

where

$$\begin{bmatrix} 1 & 0 \\ \mathbf{a} & C \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \mathbf{a}^* & C^* \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \mathbf{a} + C\mathbf{a}^* & CC^* \end{bmatrix},$$

$$i = \begin{bmatrix} 1 & 0 \\ 0 & I \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ \mathbf{a} & C \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ -C^{-1}\mathbf{a} & C^{-1} \end{bmatrix}.$$

The matrix Y can be viewed as a point in R^{pn} or, more conveniently here, as p points $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in R^n. A transformation g carries $\mathbf{1}$ and $\mathbf{y}_1, \ldots, \mathbf{y}_p$ into $\mathbf{1}$ and p vectors in $L(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p)$. The p new vectors in $L(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p)$ are generated by a matrix with positive determinant; accordingly the p new vectors with the $\mathbf{1}$-vector have the same orientation as do the p original vectors with the $\mathbf{1}$-vector.

Now suppose that $n \geq p + 1$ and that trivial observations Y with $\mathbf{1}$, $\mathbf{y}_1, \ldots, \mathbf{y}_p$ linearly dependent are excluded. Let $L^+(\mathbf{1}; \mathbf{y}_1, \ldots, \mathbf{y}_p)$ be the $(p + 1)$-dimensional subspace $L(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p)$ *together with* an orientation, the orientation of the $p + 1$ vectors $\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p$. A transformation g carries the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_p$ of a nontrivial Y into new vectors $\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_p$ in the same subspace $L^+(\mathbf{1}; \mathbf{y}_1, \ldots, \mathbf{y}_p)$ and with the same orientation relative to the $\mathbf{1}$-vector.

The definition of a transformation variable can be facilitated by notation from Section 10, Chapter Three:

$$\left[\begin{array}{c|c} 1 & 0 \\ \hline \mathbf{m}(Y) & T(Y) \end{array}\right] = \left[\begin{array}{c|cccc} 1 & 0 & \cdots & & 0 \\ \hline m_1(Y) & s_{(1)}(Y) & & & 0 \\ m_2(Y) & t_{21}(Y) & s_{(2)}(Y) & & \\ \cdot & \cdot & & \cdot & \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \\ m_p(Y) & t_{p1}(Y) & \cdots & t_{p\,p-1}(Y) & s_{(p)}(Y) \end{array}\right],$$

$$D^*(Y) = \left[\begin{array}{c} \mathbf{1}' \\ \hline \mathbf{d}_1^{*\prime}(Y) \\ \cdot \\ \cdot \\ \mathbf{d}_p^{*\prime}(Y) \end{array}\right] = \left[\begin{array}{ccc} 1 & \cdots & 1 \\ \hline d_{11}^*(Y) & \cdots & d_{1n}^*(Y) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ d_{p1}^*(Y) & \cdots & d_{pn}^*(Y) \end{array}\right] = \left[\begin{array}{c} \mathbf{1}' \\ \hline \\ \underline{D}^*(Y) \\ \\ \end{array}\right]$$

(an asterisk is added to the D-matrix of Chapter Three to distinguish it from a reference point defined in this section). The second matrix $D^*(Y)$ contains vectors $\mathbf{d}_1^*(Y), \ldots, \mathbf{d}_p^*(Y)$ obtained by successively orthogonalizing and normalizing the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in the sequence $\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p$; the first matrix contains the *coordinates* for the row vectors in Y using as a *basis* the row vectors in $D^*(Y)$.

Consider some additional notation. Let $\mathbf{y}_1^0, \ldots, \mathbf{y}_p^0$ be the projections of $(1, 0, \ldots, 0), \ldots, \pm(0, \ldots, 0, 1, 0, \ldots)$ into the subspace $L(\mathbf{1}, \mathbf{y}_1, \ldots, \mathbf{y}_p)$, the sign of the last vector being chosen so that $\mathbf{y}_1^0, \ldots, \mathbf{y}_p^0$ have the same orientation† as $\mathbf{y}_1, \ldots, \mathbf{y}_p$ in $L^+(\mathbf{1}; \mathbf{y}_1, \ldots, \mathbf{y}_p)$; let Y^0 be the corresponding matrix

$$Y^0 = \begin{bmatrix} \mathbf{1}' \\ \mathbf{y}_1^{0\prime} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_p^{0\prime} \end{bmatrix};$$

and let

$$D(Y) = D^*(Y^0).$$

The matrix $D(Y)$ contains p orthonormal vectors with appended 1-vector. The matrix $D(Y)$ depends only on the oriented subspace $L^+(\mathbf{1}; \mathbf{y}_1, \ldots, \mathbf{y}_p)$, and not otherwise on the observation Y.

Now take $D(Y)$ as the reference point on the orbit described by $L^+(\mathbf{1}; \mathbf{y}_1, \ldots, \mathbf{y}_p)$; and let $[Y]$ be the positive affine transformation that carries the row vectors of $D(Y)$ into the row vectors of Y:

$$Y = [Y]D(Y) = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline m_1(Y) & c_{11}(Y) & \cdots & c_{1p}(Y) \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ m_p(Y) & c_{p1}(Y) & \cdots & c_{pp}(Y) \end{array}\right] D(Y)$$

$$= \left[\begin{array}{c|c} 1 & 0 \\ \hline \mathbf{m}(Y) & C(Y) \end{array}\right] D(Y).$$

† Exclude for convenience of definition further trivial Y with $\mathbf{1}$, $\mathbf{y}_1^0, \ldots, \mathbf{y}_p^0$ linearly dependent.

The affine multivariate model can now be written:

$$f(E)\,dE,$$

$$Y=[Y]D(Y),\qquad D(Y)=D(E).$$

For $n \geq p+1$ the affine multivariate model† is a structural model.

The transformation $[Y]$ can be expressed in terms of triangular and orthogonal components. Let $\underset{T}{[Y]}$ be the transformation variable defined for Section 10 in Chapter Three,

$$\underset{T}{[Y]}=\begin{bmatrix}1 & 0\\ \mathbf{m}(Y) & T(Y)\end{bmatrix};$$

then

$$Y=\underset{T}{[Y]}D^*(Y).$$

And let $[Y]$ be the positive orthogonal matrix that generates the row vectors of $D^*(Y)$ from the row vectors of $D(Y)$:

$$\underset{O}{[Y]}=\begin{bmatrix}1 & 0\\ 0 & O(Y)\end{bmatrix};$$

then

$$Y=\underset{T}{[Y]}D^*(Y),\qquad D^*(Y)=\underset{O}{[Y]}D(Y),$$

$$Y=\underset{T}{[Y]}\underset{O}{[Y]}D(Y)=[Y]D(Y).$$

It follows that the transformation variable can be factored,

$$[Y]=\underset{T}{[Y]}\underset{O}{[Y]};$$

or equivalently

$$\mathbf{m}(Y)=\mathbf{m}(Y),\qquad C(Y)=T(Y)\,O(Y).$$

Note: The factorization $C(Y)=T(Y)O(Y)$ is the positive lower triangular-orthogonal factorization described in Section 6, Chapter Three.

*5 THE MULTIVARIATE MODEL: DISTRIBUTIONS

Consider the invariant differential on the response space. A transformation g applies column-by-column on the matrix Y. Its effect on the ith column is

$$[\mathbf{a}, C]\begin{bmatrix}y_{1i}\\ \cdot\\ \cdot\\ \cdot\\ y_{pi}\end{bmatrix}=\mathbf{a}+C\begin{bmatrix}y_{1i}\\ \cdot\\ \cdot\\ \cdot\\ y_{pi}\end{bmatrix},$$

† Excluding trivial points Y with linear dependence among the rows.

which has Jacobian $|C|$. Hence

$$J_{pn}(g:Y)=|C|^n=|g|^n,$$

$$J_{pn}(Y)=|C(Y)|^n=|[Y]|^n,$$

$$dm(Y)=\frac{\prod dy_{ji}}{|C(Y)|^n}=\frac{dY}{|[Y]|^n}.$$

Now consider the invariant differentials on the group;

$$\begin{bmatrix}1 & 0\\ \tilde{\mathbf{a}} & \tilde{C}\end{bmatrix}=\begin{bmatrix}1 & 0\\ \mathbf{a} & C\end{bmatrix}\begin{bmatrix}1 & 0\\ \mathbf{a}^* & C^*\end{bmatrix}.$$

The left transformation operates column-by-column. For any given column the Jacobian is $|C|$; hence

$$J=|C|^{p+1},$$

$$J(g)=|g|^{p+1},$$

$$d\mu(g)=\frac{dg}{|g|^{p+1}}.$$

The right transformation operates row-by-row. For any given row the Jacobian is $|C^*|$; hence

$$J^*=|C^*|^p,$$

$$J^*(g)=|g|^p,$$

$$d\nu(g)=\frac{dg}{|g|^p}.$$

The modular function is

$$\Delta(g)=\frac{|g|^p}{|g|^{p+1}}=\frac{1}{|g|}.$$

5.1 . General Distributions. The conditional probability element for the error position $[E]$ given the orbit $D(E)=D$ is

$$g([E]:D)\,d[E]=k(D)f([E]D)\,|[E]|^n\frac{d[E]}{|[E]|^{p+1}}$$

$$=k(D)\prod_1^n f\left(\mathbf{m}+C\begin{bmatrix}d_{1i}\\ \cdot\\ \cdot\\ \cdot\\ d_{pi}\end{bmatrix}\right)|C|^{n-p-1}\,d\mathbf{m}\,dC.$$

The structural probability element for θ given Y is

$$g^*(\theta:Y)\,d\theta = k(D(Y))f(\theta^{-1}Y)\frac{|[Y]|^n}{|\theta|^n}\frac{1}{|[Y]|}\,d\nu(\theta)$$

$$= k(D(Y))\prod_1^n f\left(\Gamma^{-1}\begin{bmatrix} y_{1i}-\mu_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{pi}-\mu_p \end{bmatrix}\right)\frac{|C(Y)|^{n-1}}{|\Gamma|^n}\frac{d\mu\,d\Gamma}{|\Gamma|^p}.$$

The error distribution provides the basis for tests of significance; the structural distribution provides the basis for general inference.

5.2 The Semidirect Decomposition. The structural equation for the affine model,

$$\mathbf{m}(Y) = \boldsymbol{\mu} + \Gamma\mathbf{m}(E),$$
$$C(Y) = \Gamma C(E),$$

can be separated into a part concerning the general level $\boldsymbol{\mu}$ and a part concerning the error scaling Γ:

$$C^{-1}(Y)(\mathbf{m}(Y)-\boldsymbol{\mu}) = C^{-1}(E)\mathbf{m}(E) = \mathbf{t}(E),$$
$$\Gamma^{-1}C(Y) = C(E).$$

The general level $\boldsymbol{\mu}$ relates to the location subgroup

$$L = \{[\mathbf{a}, I]:\quad \mathbf{a}\in R^p\},$$

and the error scaling Γ relates to the positive linear subgroup

$$S = \{[0, C]:\quad |C| > 0\}.$$

These definitions use the general location-scale notation of Problem 27 in Chapter One. The general group element can be expressed uniquely as a product, location-times-scale:

$$[\mathbf{a}, C] = [\mathbf{a}, I][0, C] = \underset{L}{[\mathbf{a}, C]}\underset{S}{[\mathbf{a}, C]},$$

or uniquely as a product, scale-times-location:

$$[\mathbf{a}, C] = [0, C][C^{-1}\mathbf{a}, I] = \underset{S}{[\mathbf{a}, C]}\underset{L}{[\mathbf{a}, C]}.$$

(See Problem 19 in Chapter Two.)

5.3 The Location Distributions. Consider the quantity $\boldsymbol{\mu}$ in relation to the positive affine group. Specification of $\boldsymbol{\mu}$ restricts the general quantity θ to the left coset of the subgroup S:

$$\{[\boldsymbol{\mu}, \Gamma]:\quad |\Gamma| > 0\} = \{[\boldsymbol{\mu}, I][0, \Gamma]:\quad |\Gamma| > 0\} = [\boldsymbol{\mu}, I]S.$$

Tests of significance and the marginal structural distribution are then based on right cosets on the error space G^*. For example, the information $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ leads to the value of the error characteristic $\mathbf{t}(E)$,

$$\mathbf{t}(E) = C^{-1}(E)\mathbf{m}(E) = C^{-1}(Y)(\mathbf{m}(Y) - \boldsymbol{\mu}_0);$$

and this value of the error characteristic $\mathbf{t}(E) = \mathbf{t}$ restricts $[\mathbf{m}(E), C(E)]$ to a *right coset*:

$$\{[\mathbf{m}, C]:\quad C^{-1}\mathbf{m} = \mathbf{t}\} = \{[0, C][C^{-1}\mathbf{m}, I]:\quad C^{-1}\mathbf{m} = \mathbf{t}\}$$
$$= S[\mathbf{t}, I].$$

See Section 6 in Chapter Two for the univariate case $p = 1$.

The full error probability distribution,

$$k(D)f([E]D)\,|C|^{n-p-1}\,d\mathbf{m}\,dC,$$

can be reexpressed by the substitution $\mathbf{m} = C\mathbf{t}$,

$$k(D)f([E]D)\,|C|^{n-p-1}\,|C|\,d\mathbf{t}\,dC,$$

and the *marginal distribution for* $\mathbf{t}$ obtained by integration:

$$g_L(\mathbf{t}:D)\,d\mathbf{t} = k(D)\int_C \prod_1^n f\left(C\begin{bmatrix} t_1 + d_{1i} \\ \cdot \\ \cdot \\ \cdot \\ t_p + d_{pi}\end{bmatrix}\right)|C|^{n-p}\,dC\cdot d\mathbf{t}.$$

The structural distribution for $\boldsymbol{\mu}$ can then be obtained by substituting $\mathbf{t} = C^{-1}(Y)(\mathbf{m}(Y) - \boldsymbol{\mu})$;

$$g_L^*(\boldsymbol{\mu}:Y)\,d\boldsymbol{\mu} = \frac{k(D(Y))}{|C(Y)|}\int_C \prod_1^n f\left(CC^{-1}(Y)\begin{bmatrix} y_{1i}-\mu_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{pi}-\mu_p\end{bmatrix}\right)|C|^{n-p}\,dC\cdot d\boldsymbol{\mu}.$$

5.4 The Scale Distributions. Consider the quantity Γ in relation to the positive affine group. Specification of Γ restricts the general quantity θ to a left coset of the subgroup L:

$$\{[\boldsymbol{\mu}, \Gamma]:\quad \boldsymbol{\mu}\in R^p\} = \{[0, \Gamma][\Gamma^{-1}\boldsymbol{\mu}, I]:\quad \boldsymbol{\mu}\in R^p]$$
$$= [0, \Gamma]L.$$

Tests of significance and the marginal structural distribution are then based on right cosets on the error space G^*. For example, the information $\Gamma = \Gamma_0$ leads to the value of the error characteristic $C(E)$,

$$C(E) = \Gamma_0^{-1}C(Y);$$

and the value of the error characteristic $C(E) = C$ restricts $[\mathrm{m}(E), C(E)]$ to a *right*† *coset:*

$$\{[\mathrm{m}, C]:C\} = \{[\mathrm{m}, I][0, C]:C\} = L[0, C].$$

The marginal distribution of C can be obtained by integration:

$$g_S(C:D)\,dC = k(D)\int_{\mathrm{m}} \prod_1^n f\left(\mathrm{m} + C\begin{bmatrix} d_{1i} \\ \vdots \\ d_{pi}\end{bmatrix}\right) d\mathrm{m}\cdot |C|^{n-p-1}\,dC.$$

The structural distribution for Γ can be obtained by the change of variable $C = \Gamma^{-1}C(Y)$ in the preceding distribution. The Jacobian evaluation for this can be avoided by using the evaluation implicit in the derivation of the full structural distribution. The marginal distribution of Γ can be obtained then by integrating out the quantity μ in the full structural distribution:

$$g_S^*(\Gamma:Y)\,d\Gamma = \frac{k(D(Y))\,|C(Y)|^{n-1}}{|\Gamma|^{n+p}}\int_\mu \prod_1^n f\left(\Gamma^{-1}\begin{bmatrix} y_{1i} - \mu_1 \\ \vdots \\ y_{pi} - \mu_p\end{bmatrix}\right) d\mu\cdot d\Gamma.$$

*6 THE MULTIVARIATE MODEL: ROTATIONAL SYMMETRY

Suppose now that the error distribution is rotationally symmetric with respect to the rotation group,

$$G_O = \left\{ h = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & o_{11} & \cdots & o_{1p} \\ \vdots & \vdots & & \vdots \\ 0 & o_{p1} & \cdots & o_{pp}\end{array}\right] = \begin{bmatrix} 1 & 0 \\ 0 & O\end{bmatrix} : \begin{array}{c} O'O = I \\ |O| = 1\end{array}\right\}:$$

$$f\left(h^{-1}\begin{bmatrix} e_1 \\ \vdots \\ e_p\end{bmatrix}\right) = f\left(\begin{bmatrix} e_1 \\ \vdots \\ e_p\end{bmatrix}\right), \qquad h \in G_O.$$

† For the location subgroup left cosets are right cosets and right cosets are left cosets: the location subgroup is a normal subgroup.

The rotation group is the stabilizer subgroup† for the error distribution.

A complementing subgroup is the location-progression group examined in Section 9 of Chapter Three:

$$G_T = \left\{ k = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ a_1 & c_1 & & & & 0 \\ a_2 & k_{21} & c_2 & & & \\ \vdots & \vdots & & \ddots & & \\ a_p & k_{p1} & \cdots & & k_{p\,p-1} & c_p\end{bmatrix} : \begin{array}{r} -\infty < a_j < \infty \\ -\infty < k_{jj'} < \infty \\ 0 < c_j < \infty \end{array}\right\}.$$

The analysis at the end of Section 4 shows that an element g of the positive affine group can be represented uniquely as a product:

$$g = kh = [g]_T[g]_O \qquad \begin{cases} k = [g]_T \in G_T, \\ h = [g]_O \in G_O. \end{cases}$$

The invariant differentials for the progression group (Section 11, Chapter Three) are

$$d\mu_T(k) = \frac{dk}{|k|_\Delta}, \qquad d\nu_T(k) = \frac{dk}{|k|_\nabla},$$

$$\Delta_T(k) = \frac{|k|_\nabla}{|k|_\Delta}.$$

The invariant differentials‡ for the rotation group are

$$d\mu_O(h) = dh = dO, \qquad d\nu_O(h) = dh = dO,$$

$$\Delta_O(h) = 1.$$

The composite differential

$$d\mu_T(k)\,d\nu_O(h)$$

for the variable kh is invariant under left multiplication by elements of G_T and right multiplication by elements of G_O. The adjusted differential

$$\frac{d\mu(kh)}{\Delta(h)}$$

† The integrability of the density can be used to show that no further symmetries are possible within the positive affine group.

‡ The differentials for an orthogonal matrix are composed of differentials on the surfaces of spheres; these differentials are invariant under orthogonal transformations; cf. Section 2.

also has these invariance properties. At the identity the component differentials $d\mu_T(k)$ and $d\nu_O(h)$ both measure Euclidean volume, one orthogonal to the other. Hence

$$d\mu(kh) = \Delta(h)\, d\mu_T(k)\, d\nu_O(h),$$

$$\frac{dg}{|g|^{p+1}} = \frac{d(kh)}{|kh|^{p+1}} = \frac{dk}{|k|_\Delta} dh.$$

The quantity θ can now be separated into an essential quantity $\underset{T}{[\theta]}$ and a redundant quantity $\underset{O}{[\theta]}$:

$$\theta = \underset{T}{[\theta]}\underset{O}{[\theta]}: \quad \begin{matrix} \underset{T}{[\theta]} \in G_T \\ \underset{O}{[\theta]} \in G_O, \end{matrix}$$

where

$$\underset{T}{[\theta]} = \left[\begin{array}{c|ccccc} 1 & 0 & 0 & \cdots & & 0 \\ \hline \mu_1 & \sigma_{(1)} & & & & 0 \\ \mu_2 & \tau_{21} & \sigma_{(2)} & & & \\ \cdot & \cdot & & \cdot & & \\ \cdot & \cdot & & & \cdot & \\ \cdot & \cdot & & & & \cdot \\ \mu_p & \tau_{p1} & \cdots & \tau_{p\,p-1} & & \sigma_{(p)} \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline \mu & \mathcal{T} \end{array}\right],$$

$$\underset{O}{[\theta]} = \left[\begin{array}{c|ccc} 1 & 0 & & 0 \\ \hline 0 & \omega_{11} & \cdots & \omega_{1p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & \omega_{p1} & \cdots & \omega_{pp} \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & \Omega \end{array}\right];$$

note that $\Gamma = \mathcal{T}\Omega$. And the structural distribution for θ can then be expressed in terms of the component quantities:

$$k(D(Y))f(\theta^{-1}Y)\frac{\|[Y]\|^{n-1}}{|\theta|^{n-1}}\, d\mu(\theta)$$

$$= k(D(Y))f(\underset{O}{[\theta]}^{-1}\underset{T}{[\theta]}^{-1}Y)\frac{\|[Y]\|^{n-1}}{|\theta|^{n-1}}\frac{d\underset{T}{[\theta]}}{|\underset{T}{[\theta]}|_\Delta}\, d\underset{O}{[\theta]}$$

$$= k(D(Y))f(\underset{T}{[\theta]}^{-1}Y)\frac{\|[Y]\|^{n-1}}{\|\underset{T}{[\theta]}\|^{n-1}}\frac{d\underset{T}{[\theta]}}{|\underset{T}{[\theta]}|_\Delta}\cdot d\underset{O}{[\theta]}.$$

The redundant quantity can then be integrated out:

$$\overset{*}{g}_T(\underset{T}{[\theta]}: Y)\, d\underset{T}{[\theta]} = k(D(Y))f(\underset{T}{[\theta]}^{-1}Y)\frac{|C(Y)|^{n-1}}{\|\underset{T}{[\theta]}\|^{n-1}}\frac{d\underset{T}{[\theta]}}{|\underset{T}{[\theta]}|_\Delta}\cdot\int d\Omega$$

$$= \prod_2^p A_j\, k(D(Y))f(\underset{T}{[\theta]}^{-1}Y)\frac{|T(Y)|^{n-1}}{\|\underset{T}{[\theta]}\|^{n-1}}\frac{d\underset{T}{[\theta]}}{|\underset{T}{[\theta]}|_\Delta};$$

the integration is performed in the pattern in Section 3: the first-column vector in Ω ranges over a unit sphere in R^p; the second conditionally over a unit sphere in R^{p-1}, . . . ; the last unit vector conditionally is determined (A_j is the area of the unit sphere in R^j).

The essential quantity $\underset{T}{[\theta]}$ can be expressed in terms of the location and scale components:

$$\underset{T}{[\theta]} = \left[\begin{array}{c|ccccc} 1 & 0 & 0 & \cdots & & 0 \\ \hline \mu_1 & \sigma_{(1)} & & & & 0 \\ \mu_2 & \tau_{21} & \sigma_{(2)} & & & \\ \cdot & \cdot & & \cdot & & \\ \cdot & \cdot & & & \cdot & \\ \cdot & \cdot & & & & \cdot \\ \mu_p & \tau_{p1} & \cdots & \tau_{p\,p-1} & & \sigma_{(p)} \end{array}\right] = \left[\begin{array}{c|c} 1 & 0 \\ \hline \mu & \mathcal{T} \end{array}\right].$$

The structural distribution can be expressed in terms of these components:

$$\prod_2^p A_j\, k(D(Y))\prod_1^n f\left(\mathcal{T}^{-1}\begin{pmatrix} y_{1i} - \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{pi} - \mu_p \end{pmatrix}\right)\frac{|T(Y)|^{n-1}}{|\mathcal{T}|^{n-1}}\frac{d\mu\, d\mathcal{T}}{|\mathcal{T}|\,|\mathcal{T}|_\Delta}.$$

*7 THE MULTIVARIATE MODEL: NORMAL ERROR

Consider the affine multivariate model with standard normal error variables:

$$f(E)\, dE = (2\pi)^{-np/2}\exp\{-\tfrac{1}{2}\sum e_{ji}^2\}\prod de_{ji},$$

$$[Y] = \theta[E], \qquad D(Y) = D(E).$$

The sum of squares in the exponential can be expressed in terms of error position (compare with Section 12 in Chapter Three):

$$\sum e_{ji}^2 = \operatorname{tr} EE' - n = \operatorname{tr} [E]DD'[E]' - n$$

$$= \operatorname{tr} [E] \begin{bmatrix} n & & & & 0 \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & 1 \end{bmatrix} [E]' - n$$

$$= \operatorname{tr} \llbracket E]\llbracket E]' - n,$$

where

$$\llbracket E] = [E] \begin{bmatrix} \sqrt{n} & & & & 0 \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{n} & 0 \\ \sqrt{n}\,\mathbf{m}(E) & C(E) \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{n} & 0 & \cdots & 0 \\ \sqrt{n}\,\bar{e}_1 & c_{11}(E) & \cdots & c_{1p}(E) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \sqrt{n}\,\bar{e}_p & c_{p1}(E) & \cdots & c_{pp}(E) \end{bmatrix}.$$

The *adjusted* position matrix $\llbracket E]$ presents the row vectors in E; it presents them relative to an orthonormal set, the row vectors of $D(E)$ with the first row vector adjusted to unit length.

The adjusted position matrix can be factored into triangular and orthogonal components as in the preceding section:

$$\llbracket E] = \underset{T}{\llbracket E]}\underset{O}{[E]},$$

where

$$\underset{T}{\llbracket E]} = \begin{bmatrix} \sqrt{n} & 0 \\ \sqrt{n}\,\mathbf{m}(E) & T(E) \end{bmatrix}$$

is the adjusted position variable of the location-progression group (compare with Section 12 in Chapter Three). The sum-of-squares can then be expressed in terms of the triangular components:

$$\sum e_{ji}^2 = \operatorname{tr} \llbracket E]\llbracket E]' - n = \operatorname{tr} \underset{T}{\llbracket E]}\underset{T}{\llbracket E]'} - n$$

$$= n \operatorname{tr} \mathbf{m}(E)\mathbf{m}'(E) + \operatorname{tr} T(E)T'(E)$$

$$= \left(\sum n\bar{e}_j^2\right) + \left(\sum t_{jj'}^2(E)\right) + \left(\sum s_{(j)}^2(E)\right).$$

7.1 General Distribution: Error. The distribution of the error position $[E]$ given the orbit is

$$g([E]:D)\,d[E] = k(D)(2\pi)^{-np/2} \exp\{-\tfrac{1}{2}(\operatorname{tr} \llbracket E]\llbracket E]' - n)\}\, |[E]|^n \frac{d[E]}{|[E]|^{p+1}}.$$

The differential can be factored by the formula in the preceding section:

$$\frac{d[E]}{|[E]|^{p+1}} = \frac{d\underset{T}{[E]}}{|\underset{T}{[E]}|_\Delta}\, d\underset{O}{[E]} = \frac{d\mathbf{m}(E)\, dT(E)}{|T(E)|\,|T(E)|_\Delta}\, dO(E).$$

The distribution of $[E]$ can then be expressed in terms of the components $\mathbf{m}(E)$, $T(E)$, $O(E)$:

$$g([E]:D)\,d[E] = k(D)(2\pi)^{-np/2} \exp\{-\tfrac{1}{2}(\operatorname{tr} \llbracket E]\llbracket E]' - n)\}$$

$$\cdot \frac{|T(E)|^n}{|T(E)|\,|T(E)|_\Delta}\, d\mathbf{m}(E)\, dT(E)\, dO(E)$$

$$= \frac{A_{n-1} \cdots A_{n-p}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2}\sum n\bar{e}_j^2 - \tfrac{1}{2}\sum t_{jj'}^2 - \tfrac{1}{2}\sum s_{(j)}^2\}$$

$$\cdot\, s_{(1)}^{n-2} \cdots s_{(p)}^{n-p-1} \prod d\sqrt{n}\,\bar{e}_j \prod dt_{jj'} \prod ds_{(j)} \frac{dO}{\prod_2^p A_j}.$$

This distribution describes a collection of independent variables: standard normal variables, chi-variables, and uniform-on-a-sphere variables. It should be noted that the triangular variable $\underset{T}{[E]}$ with components $\mathbf{m}(E)$, $T(E)$ does *not* describe the usual right cosets on the error space G^*; it describes left cosets of the orthogonal group. The distribution for $\mathbf{m}(E)$ and $T(E)$ agrees with that obtained in Section 12 of Chapter Three.

7.2 General Distribution: Structural. The structural distribution can be obtained from the general formula in Section 5.1 together with the factorization of the differential (Section 6) and the normalization constant (Section 7.1):

$$
\begin{aligned}
g^*(\theta : Y)\, d\theta &= k(D(Y)) f(\theta^{-1}Y) \frac{|[Y]|^{n-1}}{|\theta|^{n-1}}\, d\mu(\theta) \\
&= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2}(\operatorname{tr}(\theta^{-1}YY'\theta'^{-1}) - n)\} \\
&\quad \cdot \frac{n^{p/2}}{\prod_2^p A_j} \frac{|T(Y)|^{n-1}}{|\mathcal{T}|^{n-1}} \frac{d\boldsymbol{\mu}\, d\mathcal{T}\, d\Omega}{|\mathcal{T}|\,|\mathcal{T}|_\Delta}.
\end{aligned}
$$

The sum of squares in the exponents can be re-expressed:

$$
\begin{aligned}
\sum e_{ji}^2 &= \operatorname{tr} EE' - n = n \operatorname{tr} \mathbf{m}\mathbf{m}' + \operatorname{tr} CC' \\
&= n \operatorname{tr} \Gamma^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu})(\mathbf{m}(Y) - \boldsymbol{\mu})'\Gamma'^{-1} + \operatorname{tr} \Gamma^{-1}C(Y)C'(Y)\Gamma'^{-1} \\
&= n(\mathbf{m}(Y) - \boldsymbol{\mu})'(\Gamma\Gamma')^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu}) + \operatorname{tr}(\Gamma\Gamma')^{-1}C(Y)C'(Y) \\
&= n(\mathbf{m}(Y) - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu}) + \operatorname{tr}\Sigma^{-1}S(Y),
\end{aligned}
$$

where two inner-product matrices are defined by

$$\Sigma = \Gamma\Gamma' = \mathcal{T}\Omega\Omega'\mathcal{T}' = \mathcal{T}\mathcal{T}',$$

$$
\begin{aligned}
S(Y) &= C(Y)C'(Y) = T(Y)O(Y)O'(Y)T'(Y) = T(Y)T'(Y) \\
&= \begin{bmatrix} y_{11} - \bar{y}_1 & \cdots & y_{1n} - \bar{y}_1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{p1} - \bar{y}_p & \cdots & y_{pn} - \bar{y}_p \end{bmatrix} \begin{bmatrix} y_{11} - \bar{y}_1 & \cdots & y_{1n} - \bar{y}_1 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{p1} - \bar{y}_p & \cdots & y_{pn} - \bar{y}_p \end{bmatrix}'.
\end{aligned}
$$

The structural distribution is

$$
\begin{aligned}
&g^*(\theta : Y)\, d\theta \\
&= \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2}(\mathbf{m}(Y) - \boldsymbol{\mu})' n\Sigma^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu}) - \tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\} \\
&\quad \cdot \frac{n^{p/2}}{\prod_2^p A_j} \frac{|S(Y)|^{(n-1)/2}}{|\mathcal{T}|^{n-1}} \frac{d\boldsymbol{\mu}\, d\mathcal{T}\, d\Omega}{|\mathcal{T}|\,|\mathcal{T}|_\Delta} \\
&= \frac{n^{p/2}|\Sigma|^{-1/2}}{(2\pi)^{p/2}} \exp\{-\tfrac{1}{2}(\mathbf{m}(Y) - \boldsymbol{\mu})' n\Sigma^{-1}(\mathbf{m}(Y) - \boldsymbol{\mu})\}\, d\boldsymbol{\mu} \\
&\quad \cdot \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}} \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\} \frac{|S(Y)|^{(n-1)/2}}{|\mathcal{T}|^{n-1}} \frac{d\mathcal{T}}{|\mathcal{T}|_\Delta} \cdot \frac{d\Omega}{\prod_2^p A_j}.
\end{aligned}
$$

The conditional distribution of $\boldsymbol{\mu}$ given $\mathcal{T}$ is normal with mean $\mathbf{m}(Y)$ and covariance matrix $n^{-1}\Sigma = n^{-1}\mathcal{T}\mathcal{T}'$; the marginal distribution of $\mathcal{T}$ is given by the middle expression; and the distribution of Ω is uniform on the positive rotation group, and is independent of $\boldsymbol{\mu}, \mathcal{T}$.

7.3 Location Distribution: Error. The distribution of the error component

$$\mathbf{m}(E) = \begin{bmatrix} m_1(E) \\ \cdot \\ \cdot \\ \cdot \\ m_p(E) \end{bmatrix} = \begin{bmatrix} \bar{e}_1 \\ \cdot \\ \cdot \\ \cdot \\ \bar{e}_p \end{bmatrix}$$

can be obtained from the error distribution (Section 7.1):

$$\frac{1}{(2\pi)^{p/2}} \exp\{-\tfrac{1}{2}\Sigma n\bar{e}_j^2\} \Pi\, d\sqrt{n}\, \bar{e}_j.$$

It should be noted that the error component $\mathbf{m}(E)$ does *not* describe the usual right cosets on the error space G^*; it describes left cosets of the positive linear group S.

The distribution of the usual error location characteristic

$$\mathbf{t} = \mathbf{t}(E) = C^{-1}(E)\mathbf{m}(E)$$

can be obtained from the formula in Section 5.3 together with the integration properties for the distribution of $[E]$ in the normal case.

7.4 Location Distribution: Structural. The marginal structural distribution for the location quantity $\boldsymbol{\mu}$ can be obtained from the final expression in Section (7.2) by integration:

$$
\begin{aligned}
g_L^*(\boldsymbol{\mu} : Y)\, d\boldsymbol{\mu} &= \frac{n^{p/2}}{(2\pi)^{p/2}} \frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}} \\
&\quad \cdot \int_{\mathcal{T}} \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(S(Y) + n(\mathbf{m}(Y) - \boldsymbol{\mu})(\mathbf{m}(Y) - \boldsymbol{\mu})')\} \\
&\quad \cdot \frac{|S(Y)|^{(n-1)/2}}{|\mathcal{T}|^n} \frac{d\mathcal{T}}{|\mathcal{T}|_\Delta}\, d\boldsymbol{\mu} \\
&= \frac{n^{p/2}A_{n-1}\cdots A_{n-p}}{A_n\cdots A_{n-p+1}} \frac{|S(Y)|^{(n-1)/2}}{|S(Y) + n(\mathbf{m}(Y) - \boldsymbol{\mu})(\mathbf{m}(Y) - \boldsymbol{\mu})'|^{n/2}}\, d\boldsymbol{\mu} \\
&\quad \cdot \int_{\mathcal{T}} \frac{A_n\cdots A_{n-p+1}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S^*\} \frac{|S^*|^{n/2}}{|\mathcal{T}|^n} \frac{|d\mathcal{T}|}{|\mathcal{T}|_\Delta} \\
&= \frac{A_{n-p}}{A_n} \frac{n^{p/2}}{|S(Y)|^{1/2}} |I + nS^{-1}(Y)(\mathbf{m}(Y) - \boldsymbol{\mu})(\mathbf{m}(Y) - \boldsymbol{\mu})'|^{-n/2}\, d\boldsymbol{\mu} \\
&= \frac{A_{n-p}}{A_n} \frac{n^{p/2}}{|S(Y)|^{1/2}} (1 + n(\mathbf{m}(Y) - \boldsymbol{\mu})'S^{-1}(Y)(\mathbf{m}(Y) - \boldsymbol{\mu}))^{-n/2}\, d\boldsymbol{\mu};
\end{aligned}
$$

the integration property associated with the marginal distribution of $\mathcal{T}$ (Section 7.2) shows that the integral in the middle expression has value one; also check in Chapter Three (Section 8.4 and Problem 40). The marginal distribution of μ is a multivariate t-distribution but relocated and rescaled.

7.5 Scale Distribution: Error. The distribution of the error component $C(E)$ or the equivalent components $T(E)$ and $O(E)$ can be obtained from the error distribution in Section 7.1:

$$\frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\sum t_{jj'}^2-\tfrac{1}{2}\sum s_{(j)}^2\}$$
$$\cdot\, s_{(1)}^{n-2}\cdots s_{(p)}^{n-p-1}\prod dt_{jj'}\prod ds_{(j)}\cdot\frac{dO}{\prod_2^p A_j}:$$

This is a distribution describing right cosets on the error space G^*. The marginal distribution for $T(E)$, however, does not describe the usual right cosets; it describes *left* cosets of the rotation group.

7.6 Scale Distribution: Structural. The marginal structural distribution for the scale quantity Γ is directly available from the last expression in Section 7.2:

$$g_S^*(\Gamma:Y)\,d\Gamma=\frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\frac{|S(Y)|^{(n-1)/2}}{|\mathcal{T}|^{n-1}}\frac{d\mathcal{T}}{|\mathcal{T}|_\Delta}\cdot\frac{d\Omega}{\prod_2^p A_j}.$$

A related quantity of interest is the covariance matrix $\Sigma=\mathcal{T}\mathcal{T}'$. The Jacobian to obtain the distribution of Σ is available from Section 12.3 in Chapter Three:

$$\left|\frac{\partial\Sigma}{\partial\mathcal{T}}\right|=2^p\,|\mathcal{T}|_\nabla.$$

The structural distribution of Σ is

$$\frac{A_{n-1}\cdots A_{n-p}}{(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\frac{|S(Y)|^{(n-1)/2}}{|\Sigma|^{(n-1)/2}}\frac{d\Sigma}{2^p|\mathcal{T}|_\Delta|\mathcal{T}|_\nabla}$$
$$=\frac{A_{n-1}\cdots A_{n-p}}{2^p(2\pi)^{(n-1)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\frac{|S(Y)|^{(n-1)/2}}{|\Sigma|^{(n+p)/2}}\,d\Sigma.$$

NOTES AND REFERENCES

Marginal and conditional structural distributions were examined in Section 8, Chapter Two. The marginal analysis in this chapter has evolved from an analysis of these distributions for essential and redundant quantities.

The normal error distribution on the sphere was proposed by Fisher (1953) in an analysis of measurement on the sphere. As part of his analysis he derived the marginal distribution of the length $l=l(E)$. The conditional distribution of the length $l=l(E)$ given the orbit is independent of the orbit (Section 3), hence is equal to the Fisher marginal distribution. Watson and Williams (1956) used the likelihood-modulation method to simplify Fisher's derivation: The special distribution of $l=l(E)$ for $\kappa=0$ had been obtained in another context (Lord Rayleigh, 1919); the likelihood function adjusts the special distribution to give the general distribution. A survey of results for the normal distribution on the sphere is given in Stephens (1962).

The multivariate regression model has been analyzed by Fraser and M. S. Haq.

Fisher, R. A. (1953), Dispersion on a sphere, *Proc. Roy. Soc. (London)*, **A217**, 295–305.
Lord Rayleigh (1919), On the problem of random vibrations and random flights in one, two, and three dimensions, *Phil. Mag.*, (6) **37**, 321–347.
Stephens, M. A. (1962), *The statistics of directions, the Fisher and von Mises distributions*, Ph.D. thesis, University of Toronto.
Watson, G. S. and Williams, E. J. (1956), On the construction of significance tests on the circle and the sphere, *Biometrika*, **43**, 344–352.

PROBLEMS

1. Consider the composite measurement model with known scaling (Section 1), and suppose that the component error distribution is standard normal:

$$f(e_1,e_2)=\frac{1}{2\pi\sigma^2}\exp\{-\frac{1}{2\sigma^2}(e_1^2+e_2^2)\}.$$

(i) Derive the distribution of error position $[E]$; for simplicity in the final expression let

$$l^2(X)=\sum_1^n(x_{1i}-\bar{x}_1)^2+\sum_1^n(x_{2i}-\bar{x}_2)^2.$$

(ii) Derive the structural distribution for (μ_1,μ_2,φ); derive the marginal distribution of (μ_1,μ_2).

2. Consider the measurement model on the plane with unknown scaling and unknown error rotation:

$$\prod_1^n f(e_{1i},e_{2i})\prod_1^n(de_{1i}\,de_{2i}),$$

$$X=\begin{bmatrix}1 & 0 & 0\\ \mu_1 & \sigma\cos(\varphi) & -\sigma\sin(\varphi)\\ \mu_2 & \sigma\sin(\varphi) & \sigma\cos(\varphi)\end{bmatrix}E,$$

where

$$X = \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \end{pmatrix}, \qquad E = \begin{pmatrix} 1 & \cdots & 1 \\ e_{11} & \cdots & e_{1n} \\ e_{21} & \cdots & e_{2n} \end{pmatrix},$$

and $n \geq 2$. Analyze the model following the pattern and notation in Section 1.

(i) Check that the transformations form a group.
(ii) Define a transformation variable, and determine the reference point. Show that the model is a structural model. As a scale variable $s(X)$ consider $l(X)$ from Problem 1.
(iii) Derive the invariant differentials and the modular function.
(iv) Derive the distribution for error position and the structural distribution for θ.
(v) For a rotationally symmetric error distribution determine the marginal structural distribution for (μ_1, μ_2, σ).

3 (*Continuation*). Consider the preceding measurement model and suppose that the component error distribution is standard normal:

$$f(e_1, e_2) = \frac{1}{2\pi} \exp\{-\tfrac{1}{2}(e_1^2 + e_2^2)\}.$$

(i) Derive the distribution of the error position $[E]$; for simplicity in the final expression use

$$s^2(X) = \sum_1^n (x_{1i} - \bar{x}_1)^2 + \sum_1^n (x_{2i} - \bar{x}_2)^2.$$

(ii) Derive the structural distribution for $(\mu_1, \mu_2, \sigma, \varphi)$; derive the marginal distribution of (μ_1, μ_2).

4. Extend the model in Section 1 to cover measurements $(x_1, \ldots, x_p)$ on quantities $(\mu_1, \ldots, \mu_p)$:

$$f(E)\,dE = \prod_1^n f(e_{1i}, \ldots, e_{pi}) \prod_1^n (de_{1i} \cdots de_{pi}),$$
$$X = \theta E,$$

where

$$X = \begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_p' \end{pmatrix}, \quad E = \begin{pmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e}_p' \end{pmatrix}, \quad \theta = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline \mu_1 & & & \\ \vdots & & \Omega & \\ \mu_p & & & \end{array}\right);$$

the quantity Ω is an element of the rotation group on R^p, and $n \geq p$. The group of rotations on R^p is

$$G_O = \left\{ O = \begin{pmatrix} o_{11} & \cdots & o_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ o_{p1} & \cdots & o_{pp} \end{pmatrix} = (\mathbf{o}_1, \ldots, \mathbf{o}_p) : \begin{array}{c} O'O = I, \\ |O| = 1 \end{array} \right\}.$$

An element O can be described as follows: the vector $\mathbf{o}_1$ is an arbitrary unit vector in R^p and $\int d\mathbf{o}_1 = A_p$; the vector $\mathbf{o}_2$ is an arbitrary unit vector in the $(p-1)$-dimensional space orthogonal to $\mathbf{o}_1$ and $\int d\mathbf{o}_2 = A_{p-1}$ conditionally; ... ; $\mathbf{o}_{p-1}$ is an arbitrary unit vector in the 2-dimensional space orthogonal to $\mathbf{o}_1, \ldots, \mathbf{o}_{p-2}$ and $\int d\mathbf{o}_{p-1} = A_2 = 2\pi$ conditionally; $\mathbf{o}_p$ is the unique unit vector orthogonal to $\mathbf{o}_1, \ldots, \mathbf{o}_{p-1}$ such that $\mathbf{o}_1, \ldots, \mathbf{o}_p$ have the same orientation as the coordinate axes in R^p.

(i) Check that the transformations form a group, the *location-rotation group*.
(ii) Define a transformation variable and determine the reference point. Show that the model is a structural model.
(iii) Derive the invariant differentials and the modular function: Represent dO as $d\mathbf{o}_1 \cdots d\mathbf{o}_p$ subject to the constraints in the description of an element O in G.
(iv) Derive the distribution for error position and the structural distribution for θ.
(v) For a rotationally symmetric error distribution determine the marginal structural distribution for $(\mu_1, \ldots, \mu_p)$. (Compare with Problem 8 in Chapter Four.)
(vi) Suppose that the error distribution involves an additional quantity β. Obtain an expression for the marginal likelihood function.

5 (*Continuation*). Consider the preceding measurement model and suppose that the error distribution is symmetrical normal:

$$f(e_1, \ldots, e_p) = \frac{1}{(2\pi)^{p/2}\sigma^p} \exp\left\{-\frac{1}{2\sigma^2}\sum e_j^2\right\}.$$

(i) Derive the distribution of the error position $[E]$; for simplicity in the final expression let

$$l^2(X) = \sum_j \sum_i (x_{ji} - \bar{x}_j)^2.$$

Derive the marginal likelihood function for σ.
(ii) Derive the structural distribution for $\mu_1, \ldots, \mu_p, \Omega$; derive the marginal distribution of $(\mu_1, \ldots, \mu_p)$ (compare with Problem 9 in Chapter Four).

6. Extend the model in Problem 2 to cover measurements $(x_1, \ldots, x_p)$ on quantities $(\mu_1, \ldots, \mu_p)$:

$$f(E)\,dE = \prod_1^n f(e_{1i}, \ldots, e_{pi}) \prod_1^n (de_{1i} \cdots de_{pi}),$$
$$X = \theta E,$$

where

$$X = \begin{pmatrix} \mathbf{1}' \\ \mathbf{x}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_p' \end{pmatrix}, \quad E = \begin{pmatrix} \mathbf{1}' \\ \mathbf{e}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e}_p' \end{pmatrix}, \quad \theta = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline \mu_1 & & & \\ \vdots & & \sigma\Omega & \\ \mu_p & & & \end{array}\right);$$

the quantity Ω is an element of the rotation group on R^p, and $n \geq p$.

(i) Check that the transformations form a group.
(ii) Define a transformation variable and determine the reference point. Show that the model is a structural model. As a scale variable $s(X)$ consider $l(X)$ from Problem 5.

(iii) Derive the invariant differentials and the modular function.

(iv) Derive the distribution for error position and the structural distribution for θ.

(v) For a rotationally symmetric error distribution determine the marginal structural distribution for $(\mu_1, \ldots, \mu_p, \sigma)$.

(vi) Suppose the error distribution involves an additional quantity β. Obtain an expression for the marginal likelihood function.

7 (*Continuation*). Consider the preceding measurement model and suppose that the error distribution is standard normal:

$$f(e_1, \ldots, e_p) = \frac{1}{(2\pi)^{p/2}} \exp\{-\tfrac{1}{2} \sum e_j^2\}.$$

(i) Derive the distribution of the error position $[E]$; for simplicity in the final expression use

$$s^2(X) = \sum_j \sum_i (x_{ji} - \bar{x}_j)^2.$$

(ii) Derive the structural distribution for $(\mu_1, \ldots, \mu_p, \sigma, \Omega)$; derive the marginal distribution of $(\mu_1, \ldots, \mu_p)$.

8. For the following symmetrical normal distribution in R^3,

$$f(x_1, x_2, x_3 : \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\left\{-\frac{1}{2\sigma^2}((x_1 - \mu)^2 + x_2^2 + x_3^2)\right\},$$

determine the conditional distribution given that $\sum x_i^2 = 1$. Give expressions linking μ, σ of the normal distribution in R^3 with κ of the normal distribution on the sphere.

9. Consider the positive linear group

$$G = \left\{ g = \begin{bmatrix} c_{11} & \cdots & c_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ c_{p1} & \cdots & c_{pp} \end{bmatrix} : \; |g| > 0 \right\}$$

operating on points

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{p1} & \cdots & y_{pn} \end{bmatrix}$$

in Euclidean space $\mathfrak{X} = R^{pn}$ by matrix multiplication:

$$\tilde{Y} = gY.$$

(i) Check that G is a group and that the group is unitary on $\mathfrak{X}$ provided $n \geq p$ and certain trivial points are excluded.

(ii) In the pattern of Section 4 define a transformation variable $[Y] = C(Y)$ and derive the invariant differential on $\mathfrak{X}$ and the left and right invariant differentials on G.

(iii) From properties of the invariant differentials deduce the value of the Jacobian

$$\left| \frac{\partial g^{-1}}{\partial g} \right|;$$

the Jacobian for the progression group is different (compare with Problem 29 in Chapter Three).

10 (*Continuation*). Consider the *linear multivariate model* for n observations on p responses:

$$f(E)\,dE,$$
$$Y = \Gamma E,$$

where Γ is an element of the positive linear group.

(i) Derive the distribution of the error position variable $[E]$.

(ii) Derive the structural distribution for Γ.

11 (*Continuation*). Suppose the error distribution is rotationally symmetric:

$$f(O^{-1}E) = f(E)$$

for all rotations in the group

$$G_O = \left\{ O = \begin{bmatrix} o_{11} & \cdots & o_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ o_{p1} & \cdots & o_{pp} \end{bmatrix} : \; \begin{matrix} O'O = I \\ |O| = 1 \end{matrix} \right\}.$$

A complementing subgroup is the progression subgroup of positive lower-triangular matrices

$$G_T = \left\{ T = \begin{bmatrix} c_{11} & & & 0 \\ t_{21} & c_{22} & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ t_{p1} & \cdots & t_{p\,p-1} & c_{pp} \end{bmatrix} : \; \begin{matrix} 0 < c_{jj} < \infty \\ -\infty < t_{jj'} < \infty \end{matrix} \right\}.$$

Let

$$\Gamma = [\Gamma]_T[\Gamma]_O = \mathfrak{C}\Omega,$$

(i) Derive the marginal structural distribution of $\mathfrak{C}$.

(ii) Express the preceding distribution in terms of the equivalent quantity

$$\Sigma = \mathfrak{C}\mathfrak{C}' = \Gamma\Gamma'.$$

12 (*Continuation*). Consider the case of standard normal component error, and let

$$[E] = C(E) = [E]_T[E]_O = T(E)\,O(E).$$

Derive

(i) the distribution of $[E]$ given the orbit in terms of $T(E)$, $O(E)$;

(ii) the distribution of $T(E)$ given the orbit (*not a right coset distribution*);

(iii) the distribution of the error inner-product matrix $S(E) = EE' = T(E)T'(E)$ given the orbit (*not a right coset distribution*);

(iv) the structural distribution of Γ;

(v) the marginal structural distribution of $\mathfrak{C}$;

(vi) the marginal structural distribution of Σ.

13. Consider the affine multivariate model with rotational symmetry. For tests of significance the decomposition of error position is needed in the order:

$$[E] = \underset{0}{[E]}\,\underset{T}{[E]}$$

(cf. Sections 6, 7 in Chapter Two). Derive the marginal distribution of the error position $\underset{T}{[E]}$ which indexes right cosets on the error space G^*.

14 (*Continuation*). For the case of normal error derive the distribution of the error characteristic $t = t(E)$ given the orbit; see Section 7.3. This is an error distribution that corresponds to *right* cosets.

15 (*Continuation*). For the case of normal error derive the marginal distribution of the inner-product matrix for residuals

$$S(E) = T(E)T'(E)$$

(Compare with Section 12.3 in Chapter Three). This is *not* a right coset distribution.

***16.** *Regression-linear model.* Consider an error variable E

$$E = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \\ e_{11} & \cdots & e_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ e_{p1} & \cdots & e_{pn} \end{bmatrix} = \begin{bmatrix} v_1' \\ \cdot \\ \cdot \\ \cdot \\ v_r' \\ e_1' \\ \cdot \\ \cdot \\ \cdot \\ e_p' \end{bmatrix} = \begin{bmatrix} V \\ \underline{E} \end{bmatrix}$$

with error distribution

$$f(E)\,dE = f(\underline{E})\,d\_ = \prod_{i=1}^{n} f(e_{1i}, \ldots, e_{pi}) \prod_{i=1}^{n} (de_{1i}, \ldots, de_{pi}).$$

Consider a quantity

$$\theta = \begin{bmatrix} 1 & & 0 & 0 & \cdots & 0 \\ & \ddots & & \cdot & & \cdot \\ & & \ddots & \cdot & & \cdot \\ 1 & \vdots & 1 & 0 & \cdots & 0 \\ \beta_{11} & \cdots & \beta_{1r} & \gamma_{11} & \cdots & \gamma_{1p} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \beta_{p1} & \cdots & \beta_{pr} & \gamma_{p1} & \cdots & \gamma_{pp} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \mathcal{B} & \Gamma \end{bmatrix}$$

or, equivalently (general location-scale notation in Problem 27, Chapter One),

$$\underline{\theta} = [\mathcal{B}, \Gamma].$$

And consider a response matrix Y:

$$Y = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ v_{r1} & \cdots & v_{rn} \\ y_{11} & \cdots & y_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ y_{p1} & \cdots & y_{pn} \end{bmatrix} = \begin{bmatrix} v_1' \\ \cdot \\ \cdot \\ \cdot \\ v_r' \\ y_1' \\ \cdot \\ \cdot \\ \cdot \\ y_p' \end{bmatrix} = \begin{bmatrix} V \\ \underline{Y} \end{bmatrix}.$$

The *regression-linear model* is

$$f(E)\,dE, \qquad f(\underline{E})\,d\underline{E}$$

or

$$Y = \theta E, \qquad \underline{Y} = \mathcal{B}V + \Gamma\underline{E}$$

(i) Check the equivalence of the two kinds of notation.
(ii) Consider the regression-positive linear group:

$$G = \left\{ g = \begin{bmatrix} I & 0 \\ B & C \end{bmatrix} : \begin{array}{l} B \text{ is a } p \times r \text{ matrix} \\ C \text{ is a } p \times p \text{ matrix with } |C| > 0 \end{array} \right\}.$$

Check that G is a group. Describe the orbits on R^{pn} by using L^+ notation from Section 2 in Chapter 3 and Section 4 in this chapter; show that G is unitary on R^{pn} if $n \geq p + r$ and a certain degenerate set of points is deleted.

***17** (*Continuation*). Define a variable $[Y]$:

$$[Y] = \begin{bmatrix} I & 0 \\ B(Y) & C(Y) \end{bmatrix} = \underset{T}{[Y]}\,\underset{0}{[Y]} = \begin{bmatrix} I & 0 \\ B(Y) & T(Y) \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & O(Y) \end{bmatrix},$$

and a point $D(Y)$ in R^{pn}:

$$D(Y) = \begin{bmatrix} v_1' \\ \cdot \\ \cdot \\ \cdot \\ v_r' \\ d_1' \\ \cdot \\ \cdot \\ \cdot \\ d_p' \end{bmatrix} = \begin{bmatrix} V \\ \underline{D}(Y) \end{bmatrix}.$$

Note: Intermediate $D^*(Y)$ and Y^0 can be defined here from Problem 33 in Chapter Three in the same way as corresponding matrices in Section 4 were defined from Section 10 in Chapter Three. Show that $[Y]$ is a transformation variable and $D(Y)$ is a reference point. Check the alternative notation: $[B(Y), C(Y)]$ and $\underline{Y} = B(Y)V + C(Y)\,\underline{D}(Y)$.

*18 (*Continuation*). Verify the following invariant differentials:

$$dm(Y) = \frac{dY}{|[Y]|^n} = \frac{dY}{|C(Y)|^n},$$

$$d\mu(g) = \frac{dg}{|g|^{p+r}} = \frac{dB\,dC}{|C|^{p+r}},$$

$$d\nu(g) = \frac{dg}{|g|^p} = \frac{dB\,dC}{|C|^p},$$

$$\Delta(g) = \frac{1}{|g|^r} = \frac{1}{|C|^r}.$$

*19 (*Continuation*). Derive the following distributions:

$$g([E]:D)\,d[E] = k(D) f([E]D)\,|[E]|^n \frac{d[E]}{|[E]|^{p+r}}$$

$$= k(D) \prod_1^n f\left(B\begin{bmatrix} v_{1i} \\ \cdot \\ \cdot \\ \cdot \\ v_{ri}\end{bmatrix} + C\begin{bmatrix} d_{1i} \\ \cdot \\ \cdot \\ \cdot \\ d_{pi}\end{bmatrix}\right) |C|^{n-p-r}\,dB\,dC,$$

$$g^*(\theta:Y)\,d\theta = k(D) f(\theta^{-1}Y)\frac{|[Y]|^n}{|\theta|^n}\frac{1}{|[Y]|^r}\,d\nu(\theta)$$

$$= k(D) \prod_1^n f\left(\Gamma^{-1}\left(\begin{bmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi}\end{bmatrix} - \mathcal{B}\begin{bmatrix} v_{1i} \\ \cdot \\ \cdot \\ \cdot \\ v_{ri}\end{bmatrix}\right)\right) \frac{|C(Y)|^{n-r}}{|\Gamma|^n}\frac{d\mathcal{B}\,d\Gamma}{|\Gamma|^p}.$$

*20 (*Continuation*). (i) Derive the right-coset location distribution $g_L(H:D)\,dH$ for the error variable $H = C^{-1}B$; derive the structural distribution for $g_L^*(\mathcal{B}:Y)\,d\mathcal{B}$ for $\mathcal{B}$.

(ii) Derive the right-coset scale distribution $g_S(C:D)\,dC$ for the error variable $C = C(E)$; derive the structural distribution $g_S^*(\Gamma:Y)\,d\Gamma$ for Γ (cf. Problem 36, Chapter Three).

*21 (*Continuation*). Suppose the error distribution is rotationally symmetric:

$$f(h^{-1}E) = f(E)$$

for all rotations in the group

$$G_O = \left\{h = \begin{bmatrix} I & 0 \\ 0 & O\end{bmatrix} : \quad \begin{matrix} O'O = I \\ |O| = 1\end{matrix}\right\}.$$

A complementing subgroup is the regression-progression group in Problem 32 in Chapter Three,

$$G_T = \left\{k = \begin{bmatrix} I & 0 \\ B & T\end{bmatrix} : \quad \begin{matrix} B \text{ is a } p \times r \text{ matrix} \\ T \text{ is a } p \times p \text{ positive-lower-triangular}\end{matrix}\right\}.$$

The invariant differentials for G_O are given in Section 6; the invariant differentials for G_T are given in Problem 34, Chapter Three. Let

$$\theta = \underset{T}{[\theta]}\underset{O}{[\theta]} = \begin{bmatrix} I & 0 \\ \mathcal{B} & \mathcal{C}\end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & \Omega\end{bmatrix}.$$

(i) Derive the marginal structural distribution of $(\mathcal{B}, \mathcal{C})$.

(ii) Express the preceding structural distribution in terms of the equivalent quantity $(\mathcal{B}, \Sigma)$, where

$$\Sigma = \Gamma\Gamma' = \mathcal{C}\mathcal{C}'.$$

*22 (*Continuation*). Consider the case of standard normal component error. Let

$$[E] = \underset{T}{[E]}\underset{O}{[E]} = \begin{bmatrix} I & 0 \\ B(E) & T(E)\end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & O(E)\end{bmatrix},$$

where $C(E) = T(E)\,O(E)$. Note that $\underline{E}\underline{E}' = (BV + C\underline{D})(BV + C\underline{D})'$

(i) Derive the distribution of $[E]$, given orbit in terms of $B(E)$, $T(E)$, $O(E)$. Record the marginal distribution of $B(E)$, $T(E)$, given orbit (*not a right coset distribution*).

(ii) Derive the distribution of the error inner-product matrix for residuals

$$S(E) = C(E)\,C'(E) = T(E)\,T'(E)$$

(*not a right coset distribution*). Compare with the distribution in Section 12.3 in Chapter Three and in Problem 15 in this chapter.

(iii) Derive the structural distribution for $\mathcal{B}$, $\mathcal{C}$, Ω.

(iv) Determine the marginal structural distribution of $\mathcal{B}$ and the marginal structural distribution of $\mathcal{C}$.

(v) Determine the marginal structural distribution of $\Sigma = \Gamma\Gamma' = \mathcal{C}\mathcal{C}'$.

*23 (*Continuation: case of normal error*). Derive the distribution of the error characteristic

$$H = H(E) = C^{-1}(E)\,B(E),$$

an analog of the t-variable in Section 7.3 and Problem 14. This is a *right-coset* distribution of the "general location" subgroup, the regression subgroup; it is the distribution appropriate to tests of significance concerning location and to general inference concerning the quantity $\mathcal{B}$.

CHAPTER SIX

Local Analysis

The structural models in preceding chapters are designed to describe systems in which the primary quantity θ is a transformation in a group, a transformation that carries an error value from within the system into a response value on the chosen measurement scales. A change in the quantity θ is a change in the transformation, and a change in the transformation produces a change in the response value.

In some systems a weaker condition exists. A change in the quantity θ is a change in a transformation carrying internal error into a response value, and a change in the transformation produces a change in the response value. The pattern of change in the response values, however, may be different near one θ value than near another θ value; in other words, the transformations may *not produce a group of transformations on the possible response values.*

In this chapter the weaker condition is examined for a real variable and a real quantity. An increase in the quantity θ is assumed to cause an increase in the response variable x. But the pattern of increase may be different near different θ values.

1 THE STOCHASTICALLY MONOTONE MODEL

Consider a real-valued quantity θ and a real-valued response x. Suppose that an increase in θ produces an increase in x, and let the classical model for x be given by the distribution function $F(x:\theta)$.

Suppose $f(x:\theta) = \partial F(x:\theta)/\partial x$ is continuously differentiable with respect to x and θ. The total differential for F at θ_0 is

$$dF = \frac{\partial}{\partial x} F(x:\theta_0)\,dx + \left[\frac{\partial}{\partial \theta} F(x:\theta)\right]_{\theta=\theta_0} d\theta = F_x(x:\theta_0)\,dx + F_\theta(x:\theta_0)\,d\theta.$$

With the probability level F held constant the differential dx that corresponds to the differential $d\theta$ at θ_0 is

$$dx = -\frac{F_\theta(x:\theta_0)}{F_x(x:\theta_0)}\,d\theta.$$

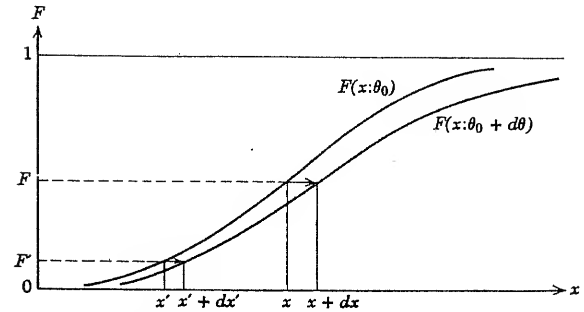


Figure 1 A change $d\theta$ at θ_0; the corresponding change in the response variable at various points x (at various values for F).

Suppose that this differential dx is not equal to $d\theta$ at some or all values of F. Then let† $l(x, \theta_0)$ be a *linearized variable*, an increasing function of x that has differential dl *uniformly equal to* $d\theta$ at θ_0:

$$dl = -\frac{F_x(x:\theta_0)}{F_\theta(x:\theta_0)}\,dx = d\theta$$

$$l(x, \theta_0) = \int^x -\frac{F_x(t:\theta_0)}{F_\theta(t:\theta_0)}\,dt$$

(See Figure 1.) The linearized variable is determined except for an additive constant corresponding to the absent lower limit of integration.

Let $H(l:\theta)$ be the distribution function for the new variable $l(x, \theta_0)$:

$$H(l:\theta) = F(x(l, \theta_0):\theta),$$

where $x(l, \theta_0)$ is the inverse function obtained by solving

$$l(x, \theta_0) = l$$

for x.

For the new variable l the differential dl that corresponds to the differential $d\theta$ at θ_0 is

$$dl = -\frac{H_\theta(l:\theta_0)}{H_l(l:\theta_0)}\,d\theta = -\frac{F_\theta(x:\theta_0)}{F_x(x:\theta_0)\,\dfrac{dx(l, \theta_0)}{dl}}\,d\theta = d\theta.$$

This checks that the new variable $l(x, \theta_0)$ has the required property. Note that $H_\theta(l:\theta_0) = -H_l(l:\theta_0)$ for all l.

† The linearized variable $l(x, \theta)$ here should not be confused with the log-likelihood function $l(x:\theta)$ in Chapter Four.

The model for the new variable can be expressed as

$$H(l - (\theta - \theta_0):\theta_0)$$

to a first derivative approximation for θ near θ_0:

$$H(l - (\theta_0 - \theta_0):\theta_0) = H(l:\theta_0),$$

$$\left[\frac{\partial}{\partial\theta} H(l - (\theta - \theta_0):\theta_0)\right]_{\theta=\theta_0} = -H_l(l:\theta_0) = H_\theta(l:\theta_0).$$

The model, to a first derivative approximation at θ_0, can then be viewed as the classical model based on the following simple measurement model

$$H(e:\theta_0),$$
$$l = [(\theta - \theta_0), 1]e.$$

The assumptions at the beginning of the chapter present a change in θ as a change in a transformation applied to internal error. The preceding simple measurement model is then the appropriate model for inference near θ_0.

Now consider a sequence $(x_1, \ldots, x_n)$ of response values. The model for the response sequence is

$$\prod_1^n F(x_i:\theta);$$

or, in terms of the transformed response sequence, $(l_1, \ldots, l_n) = (l(x_1, \theta), \ldots, l(x_n, \theta))$, the model is

$$\prod_1^n H(l_i:\theta).$$

The related simple measurement model uses a position variable $r(\mathbf{l})$ and an orbital variable

$$\mathbf{d}(\mathbf{l}) = (l_1 - r(\mathbf{l}), \ldots, l_n - r(\mathbf{l}))':$$

A simple choice is $r(\mathbf{l}) = l_1$ and

$$\mathbf{d}(\mathbf{l}) = (0, l_2 - l_1, \ldots, l_n - l_1).$$

The marginal probability element for d is

$$\int_{-\infty}^{\infty} \prod_1^n H_l(t + d_i:\theta)\, dt \cdot dd_2 \cdots dd_n.$$

A first derivative change in θ at θ_0 should shift the distribution of $\mathbf{l}$ along the orbits; accordingly the marginal distribution of $\mathbf{d}$ should have a derivative

equal to zero at θ_0:

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\prod_1^n H_l(t + d_i:\theta)\, dt\Big|_{\theta=\theta_0} = \int_{-\infty}^{\infty}\prod_1^n H_l(t + d_i:\theta_0)\sum_1^n \frac{H_{l\theta}(t + d_i:\theta_0)}{H_l(t + d_i:\theta_0)}\, dt$$

$$= -\int_{-\infty}^{\infty}\prod_1^n H_l(t + d_i:\theta_0)\sum_1^n \frac{H_{ll}(t + d_i:\theta_0)}{H_l(t + d_i:\theta_0)}\, dt$$

$$= -\int_{-\infty}^{\infty}\frac{\partial}{\partial t}\prod_1^n H_l(t + d_i:\theta_0)\, dt$$

$$= -\left[\prod_1^n H_l(t + d_i:\theta_0)\right]_{-\infty}^{\infty} = 0.$$

The reduction uses $H_{ll}(l:\theta_0) = -H_{\theta l}(l:\theta_0)$, a consequence of $H_l(l:\theta_0) = -H_\theta(l:\theta_0)$.

The conditional distribution for the location variable l_1 is

$$g(l_1:\mathbf{d}, \theta)\, dl_1 = \frac{\prod_1^n H_l(l_1 + d_i:\theta)}{\int_{-\infty}^{\infty}\prod_1^n H_l(t + d_i:\theta)\, dt}\, dl_1.$$

The conditional distribution should have the same linearized property at θ_0 as has the distribution of a single l; this is easily checked by showing that

$$\left[\frac{\partial}{\partial\theta} g(l_1:\mathbf{d}, \theta)\right]_{\theta=\theta_0} = -\frac{\partial}{\partial l_1} g(l_1:\mathbf{d}, \theta_0)$$

(see Problem 1).

The conditional model can be expressed as

$$g(l_1 - (\theta - \theta_0):\mathbf{d}, \theta_0)$$

to a first derivative approximation for θ near θ_0. The model can then be viewed as the classical model derived from the following reduced simple measurement model

$$g(e_1:\mathbf{d}(\mathbf{l}), \theta_0)$$
$$l_1 = (\theta - \theta_0) + e_1.$$

The assumptions at the beginning of the chapter then imply that the reduced model is the appropriate model for inference near θ_0.

The structural probability element for θ at θ_0 is

$$g(l_1:\mathbf{d}(l), \theta_0)\, d\theta = \frac{\prod_1^n H_i(l_1 + d_i:\theta_0)}{\int_{-\infty}^{\infty} \prod_1^n H_i(t + d_i:\theta_0)\, dt}\, d\theta$$

$$= \frac{\prod_1^n F_\theta(x_i:\theta_0)}{\int_{-\infty}^{\infty} \prod_1^n F_\theta(x(t + d_i, \theta_0):\theta_0)\, dt}\, d\theta.$$

This element of probability at θ_0 uses the linearized variable $l(x, \theta_0)$ for that θ value.

Now consider the marginal likelihood function based on the orbital variable. The marginal element for $\mathbf{d}$ is

$$\int_{-\infty}^{\infty} \prod_1^n H_i(t + d_i:\theta_0)\, dt \cdot dd_2 \cdots dd_n$$

$$= \int_{-\infty}^{\infty} (-1)^n \prod_1^n F_\theta(x(t + d_i, \theta_0):\theta_0)\, dt\, \frac{d\mathbf{l}}{dl_1}$$

$$= \frac{\int_{-\infty}^{\infty} \prod_1^n F_\theta\big(x(t + l(x_i, \theta_0) - l(x_1, \theta_0), \theta_0):\theta_0\big)\, dt}{\prod \dfrac{F_\theta(x_i:\theta_0)}{F_x(x_i:\theta_0)}} \cdot \frac{\prod dx_i}{dl_1}.$$

The differential dl_1 for the position variable on the orbit can be related to differential length ds along the inverse image of the orbit—on the space of the x's:

$$(ds)^2 = \sum_1^n (dx_i)^2 = \sum_1^n \left(\frac{dx_i(l_1 + d_i, \theta_0)}{dl_1}\right)^2 (dl_1)^2$$

$$= \sum_1^n \left(\frac{F_\theta(x_i:\theta_0)}{F_x(x_i:\theta_0)}\right)^2 (dl_1)^2.$$

The marginal element for the orbital variable can now be expressed in terms of Euclidean volume, on the space of the x's, cross-sectional to the inverse image of the orbit. This gives the *marginal likelihood function for* θ:

$$\int_{-\infty}^{\infty} \prod_1^n F_\theta\big(x(t + l(x_i, \theta) - l(x_1, \theta), \theta):\theta\big)\, dt \cdot \frac{\left[\sum \left(\dfrac{F_\theta(x_i:\theta)}{F_x(x_i:\theta)}\right)^2\right]^{1/2}}{\prod \dfrac{F_\theta(x_i:\theta)}{F_x(x_i:\theta)}}$$

The marginal likelihood function gives secondary information concerning θ, information derived from the *orbit* at the observed response vector.

2 THE LINEARIZED POISSON

The Poisson model is used to describe the frequency of an event that can occur randomly in an interval of space or time. The quantity θ is the mean frequency for the interval of space or time. An increase in θ corresponds to a compression of the process and a consequent stochastic increase in the frequency. These are the necessary ingredients for the application of the methods in the preceding section, except for the *discreteness* of the distribution.

If the quantity θ is large, however, the distribution for the Poisson variable spreads over a broad range of integers and is closely approximated by a continuous distribution. This section applies the methods in the preceding section to the approximating continuous distribution.

The Poisson distribution function

$$\sum_0^x \frac{\theta^x}{x!} \exp\{-\theta\}$$

can be viewed as giving the cumulative probability to the point $x + \frac{1}{2}$ for the approximating continuous distribution $F(x:\theta)$:

$$F(x + \tfrac{1}{2}:\theta) = \sum_0^x \frac{\theta^x}{x!} \exp\{-\theta\}.$$

An alternative expression for F can be obtained by differentiating with respect to θ and then integrating back:

$$\frac{\partial}{\partial\theta} F(x + \tfrac{1}{2}:\theta) = \sum_0^x -\frac{\theta^x}{x!} \exp\{-\theta\} + \sum_1^x \frac{\theta^{x-1}}{(x-1)!} \exp\{-\theta\}$$

$$= -\frac{\theta^x}{x!} \exp\{-\theta\}$$

$$F(x + \tfrac{1}{2}:\theta) = \int_\theta^\infty \frac{\theta^x}{x!} \exp\{-\theta\}\, d\theta = \int_\theta^\infty \frac{\theta^x}{\Gamma(x+1)} \exp\{-\theta\}\, d\theta.$$

In this alternative form the distribution function extends smoothly and continuously for all values of x greater than $x = -\frac{1}{2}$:

$$F(x:\theta) = \frac{\int_\theta^\infty t^{x-1/2} \exp\{-t\}\, dt}{\Gamma(x + \frac{1}{2})}$$

$$= \frac{\int_\theta^\infty t^{x-1/2} \exp\{-t\}\, dt}{\int_0^\infty t^{x-1/2} \exp\{-t\}\, dt}.$$

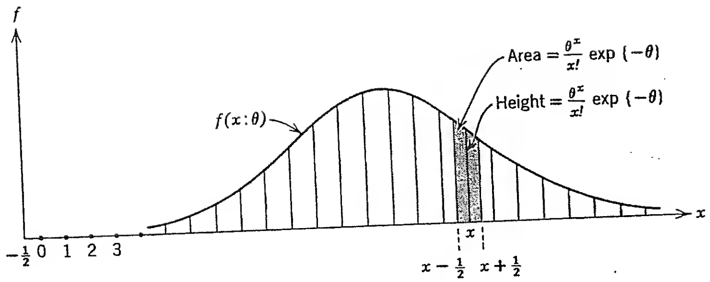


Figure 2. The ordinary Poisson probability function is given by the vertical bars. The continuous probability function is designated $f(x:\theta)$. The continuous Poisson density does not pass precisely through the tips of the bars; rather a bar gives the average height of $f(x:\theta)$ on the unit interval centered at the base point of the bar.

This distribution on the range $(-\frac{1}{2}, \infty)$ will be called the *continuous Poisson distribution*. The probability for the continuous Poisson between $x - \frac{1}{2}$ and $x + \frac{1}{2}$ for x an integer is equal to the probability at x for the ordinary Poisson distribution (see Figure 2 and Problem 5).

Now consider the linearized variable $l(x, \theta_0)$ for the continuous Poisson distribution. The derivative with respect to θ is available from the preceding analysis:

$$F_\theta(x:\theta) = -\frac{\theta^{x-\frac{1}{2}}}{\Gamma(x+\frac{1}{2})}\exp\{-\theta\}.$$

The derivative with respect to x seems unavailable explicitly. A series expansion can be calculated with considerable difficulty; a first-order approximation relative to $(1/\theta)$ is, however, available almost trivially: the density function $f(x:\theta) = F_x(x:\theta)$ is closely approximated by the Poisson probability bars (see Figure 2). This gives the approximation

$$F_x(x:\theta) \approx \frac{\theta^x}{\Gamma(x+1)}\exp\{-\theta\}.$$

The integrand for the linearizing transformation is

$$-\frac{F_x(x:\theta_0)}{F_\theta(x:\theta_0)} = \frac{\dfrac{\theta_0^x}{\Gamma(x+1)}\exp\{-\theta_0\}}{\dfrac{\theta_0^{x-\frac{1}{2}}}{\Gamma(x+\frac{1}{2})}\exp\{-\theta_0\}}$$

$$= \theta_0^{\frac{1}{2}}\frac{\Gamma(x+\frac{1}{2})}{\Gamma(x+1)}.$$

Stirling's formula,

$$\Gamma(x+1) = \sqrt{2\pi}\,x^{x+\frac{1}{2}}\exp\left\{-x + \frac{1}{12x} - \frac{1}{360x^3} + \cdots\right\}$$

can be used to simplify the ratio of gamma functions:

$$-\frac{F_x(x:\theta_0)}{F_\theta(x:\theta_0)} \approx \theta_0^{\frac{1}{2}}\frac{(x-\frac{1}{2})^x\exp\{-x+\frac{1}{2}\}}{x^{x+\frac{1}{2}}\exp\{-x\}}$$

$$= \left(\frac{\theta_0}{x}\right)^{\frac{1}{2}}(1-\tfrac{1}{2}x^{-1})^x\exp\{\tfrac{1}{2}\}$$

$$\approx \left(\frac{\theta_0}{x}\right)^{\frac{1}{2}}.$$

The approximation applies for x and θ_0 large. The linearized variable is

$$l(x, \theta_0) = \int^x\left(\frac{\theta_0}{t}\right)^{\frac{1}{2}}dt = 2\theta_0^{\frac{1}{2}}x^{\frac{1}{2}}.$$

The form of the linearized variable suggests linearization with respect to the modified quantity $\sqrt{\theta}$:

$$-\frac{F_x(x:\theta_0)}{F_{\sqrt{\theta}}(x:\theta_0)} \approx \left(\frac{\theta_0}{x}\right)^{\frac{1}{2}}\frac{1}{[d\theta/d\sqrt{\theta}]_{\theta=\theta_0}} = \frac{1}{2\sqrt{x}}.$$

The linearized variable relative to the quantity $\sqrt{\theta}$ at $\sqrt{\theta_0}$ is

$$l(x, \sqrt{\theta_0}) = \int^x\frac{1}{2\sqrt{t}}\,dt = \sqrt{x}.$$

This linearized variable does not depend on θ_0; it is a linearized variable that applies generally, provided, of course, that the quantity θ is large. It follows then that the distribution of the error variable

$$e = \sqrt{x} - \sqrt{\theta}$$

does not depend on θ, provided the quantity θ is large. The transformation $\sqrt{x}$ can be referred to as a *distribution-stabilizing transformation*.

IN SUMMARY: *The Poisson model in a frequency application can be approximated for large values of θ by the simple measurement model*

$$e,$$

$$\sqrt{x} = \sqrt{\theta} + e.$$

The model has an error variable e with an approximate distribution (the limiting form is examined in the next section); *and the model has a measurement $\sqrt{x}$, and a location quantity $\sqrt{\theta}$.*

3 DISTRIBUTION OF THE LINEARIZED POISSON

For a Poisson variable x consider the distribution of the error variable

$$e = \sqrt{x} - \sqrt{\theta}$$

for large θ. This can be examined most easily by treating x as a continuous Poisson variable and using the ordinary Poisson probability function as the approximating density function:

$$\begin{aligned} f(x:\theta) &= \frac{\theta^x}{x!} \exp\{-\theta\}\, dx \\ &= \frac{\theta^x}{\Gamma(x+1)} \exp\{-\theta\} 2\sqrt{x}\, d\sqrt{x} \\ &= \sqrt{\frac{2}{\pi}} \left(\frac{\theta}{x}\right)^x \exp\left\{x - \theta - \frac{1}{12x} + \cdots\right\} d\sqrt{x}. \end{aligned}$$

Let $g(e:\theta)\, de$ be the probability element for the error variable e. Then

$$\begin{aligned} \ln g(e:\theta) &= \ln\sqrt{\frac{2}{\pi}} + \left(x - \theta - \frac{1}{12x} + \cdots\right) - 2x(\ln\sqrt{x} - \ln\sqrt{\theta}) \\ &= \ln\sqrt{\frac{2}{\pi}} + \left((\sqrt{\theta}+e)^2 - \theta - \frac{1}{12\theta}\left(1 + \frac{e}{\sqrt{\theta}}\right)^{-2} + \cdots\right) \\ &\quad - 2(e+\sqrt{\theta})^2 \ln\left(1 + \frac{e}{\sqrt{\theta}}\right) \\ &= \ln\sqrt{\frac{2}{\pi}} + e^2 + 2e\sqrt{\theta} - \frac{1}{12\theta}\left\{1 - 2\frac{e}{\sqrt{\theta}} + \frac{2.3}{2.1}\frac{e^2}{\theta} - \cdots\right\} \\ &\quad - 2(e^2 + 2e\sqrt{\theta} + \theta)\left\{\frac{e}{\sqrt{\theta}} - \frac{e^2}{2\theta} + \frac{e^3}{3\theta^{3/2}} - \frac{e^4}{4\theta^2} + \cdots\right\} \\ &= \ln\sqrt{\frac{2}{\pi}} + e^2 + 2e\sqrt{\theta} - \frac{1}{12\theta} + \cdots \\ &\quad - 2\frac{e^3}{\sqrt{\theta}} + \frac{e^4}{\theta} - \cdots \\ &\quad - 4e^2 + 2\frac{e^3}{\sqrt{\theta}} - \frac{4}{3}\frac{e^4}{\theta} + \cdots \\ &\quad - 2e\sqrt{\theta} + e^2 - \frac{2}{3}\frac{e^3}{\sqrt{\theta}} + \frac{1}{2}\frac{e^4}{\theta} - \cdots \\ &= \ln\sqrt{\frac{2}{\pi}} - 2e^2 - \frac{2}{3}\frac{e^3}{\sqrt{\theta}} - \frac{1-2e^4}{12\theta} \pm \cdots. \end{aligned}$$

Hence

$$g(e:\theta)\, de = \sqrt{\frac{2}{\pi}} \exp\{-2e^2\} \exp\left\{-\frac{2}{3}\frac{e^3}{\sqrt{\theta}} - \frac{1-2e^4}{12\theta} \pm \cdots\right\}.$$

The limiting form of the distribution of the error variable e is normal with mean 0 and standard deviation $\frac{1}{2}$. The final factor can be expressed in terms of power series in $(e/\sqrt{\theta})$; it describes the departure from the limiting form of the stabilized distribution.

The series involved in the preceding analysis are power series that converge for $|e| < \sqrt{\theta}$. It follows that the log-density converges uniformly in any finite range for e as $\theta \to \infty$. The Poisson and normal densities decrease monotonely about their maximum points; hence the distribution of the error variable of the stabilized Poisson approaches the normal density uniformly as $\theta \to \infty$.

The Poisson distribution is usually approximated by using the variable

$$z = \frac{x - \theta}{\sqrt{\theta}}$$

as a standard normal variable. The limiting form of the distribution of z can now be examined. The Poisson probability function treated as a density function (Section 2) is

$$\frac{\theta^x}{x!} \exp\{-\theta\}\, dx = \frac{1}{(2\pi)^{1/2}} \frac{\theta^x}{x^{x+1/2}} \exp\left\{x - \theta - \frac{1}{12x} + \cdots\right\} d\sqrt{\theta}\, z.$$

Let $h(z:\theta)$ be the corresponding density function for z; then

$$\begin{aligned} \ln h(z:\theta) &= \ln\sqrt{\frac{1}{2\pi}} + \left(x - \theta - \frac{1}{12x} + \cdots\right) - (x + \tfrac{1}{2})(\ln x - \ln\theta) \\ &= \ln\sqrt{\frac{1}{2\pi}} + z\sqrt{\theta} - \frac{1}{12\theta}\left(1 + \frac{z}{\sqrt{\theta}}\right)^{-1} - (\theta + z\sqrt{\theta} + \tfrac{1}{2}) \ln\left(1 + \frac{z}{\sqrt{\theta}}\right) \\ &= \ln\sqrt{\frac{1}{2\pi}} + z\sqrt{\theta} - \frac{1}{12\theta} + \cdots \\ &\quad - (\theta + z\sqrt{\theta} + \tfrac{1}{2})\left(\frac{z}{\sqrt{\theta}} - \frac{z^2}{2\theta} + \frac{z^3}{3\theta^{3/2}} - \frac{z^4}{4\theta^2} + \cdots\right) \\ &= \ln\sqrt{\frac{1}{2\pi}} + z\sqrt{\theta} - \frac{1}{12\theta} + \cdots \\ &\quad - z\sqrt{\theta} + \frac{z^2}{2} - \frac{z^3}{3\sqrt{\theta}} + \frac{z^4}{4\theta} - \cdots \\ &\quad - z^2 + \frac{z^3}{2\sqrt{\theta}} - \frac{z^4}{3\theta} + \cdots \\ &\quad - \frac{1}{2}\frac{z}{\sqrt{\theta}} + \frac{z^2}{4\theta} - \cdots \\ &= \ln\sqrt{\frac{1}{2\pi}} - \frac{z^2}{2} + \frac{z^3 - 3z}{6\sqrt{\theta}} + \frac{3z^2 - 1 - z^4}{12\theta} \pm \cdots \end{aligned}$$

Hence

$$h(z:\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \exp\left\{\frac{z^3 - 3z}{6\sqrt{\theta}} + \frac{3z^2 - 1 - z^4}{12\theta} \pm \cdots\right\}.$$

The variable z has a limiting normal form with mean 0 and standard deviation 1.

The variable e has limiting *error* form,

$$e = \sqrt{x} - \sqrt{\theta},$$

as $\theta \to \infty$; the variable z,

$$z = \frac{x}{\sqrt{\theta}} - \sqrt{\theta},$$

does not.

Both variables, e and z, have a limiting normal distribution. It is perhaps of interest to compare the approach of these distributions to the limiting normal form. The first variable can be given the same standard deviation as the second variable z by introducing a factor 2:

$$e^* = 2e = 2\sqrt{x} - 2\sqrt{\theta}.$$

Now both variables, e^* and z, have limiting standard normal distributions; their densities are

$$\tfrac{1}{2}g\left(\frac{e^*}{2}:\theta\right) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{e^{*2}}{2}\right\} \exp\left\{-\frac{e^{*3}}{12\sqrt{\theta}} - \frac{1 - e^{*4}/8}{12\theta} \pm \cdots\right\},$$

$$h(z:\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \exp\left\{-\frac{-2z^3 + 6z}{12\sqrt{\theta}} - \frac{1 - 3z^2 + z^4}{12\theta} \cdots\right\}.$$

The first-order correction term for e^* is considerably smaller than that for z except for a small range between $1\frac{1}{2}$ and 2 standard deviations from the center of the distribution. The second-order correction for e^* seems more modest than that for z. Both the linearization with respect to the quantity and the rate of approach to limiting normality favor the error variable

$$e = \sqrt{x} - \sqrt{\theta}$$

with an approximating normal distribution with mean 0 and standard deviation $\frac{1}{2}$.

Consider an example illustrating the approximating measurement model. With observation $x = 121$ from a Poisson process the approximating model is

$$\sqrt{\frac{2}{\pi}} \exp\{-2e^2\}\, de,$$

$$11 = \sqrt{\theta} + e.$$

A 95% probability interval for e is $(-0.98, +0.98)$; a 95% structural probability interval for $\sqrt{\theta}$ is 11 ± 0.98, and for θ it is $(100.4, 143.5)$.

NOTES AND REFERENCES

The methods of local analysis in Section 1 were developed in Fraser (1964a,b). Some related aspects of invariance are discussed in Brillinger (1963) and Fraser (1967).

The distribution function of the continuous Poisson distribution is a ratio of an incomplete to a complete gamma function. Its form as the continuous Poisson was derived by Y. S. Lee. An asymptotic expression for the linearizing transformation at θ_0 has been derived by G. Van Belle.

Brillinger, D. R. (1963), Necessary and sufficient conditions for a statistical model to be invariant under a Lie group, *Ann. Math. Statistics*, **34**, 492–500.

Fraser, D. A. S. (1964a), Local conditional sufficiency, *J. Roy. Statist. Soc.*, **B26**, 52–62.

Fraser, D. A. S. (1964b), On local inference and information, *J. Roy. Statist. Soc.*, **B26**, 253–260.

Fraser, D. A. S. (1967), Statistical models and invariance, *Ann. Math. Statistics*, **38**, 1061–1067.

Van Belle, G. (1967), *Location analysis and the Poisson distribution*, Ph.D. thesis, University of Toronto.

PROBLEMS

1. Show that the conditional distribution of l_1 given the orbit $\mathbf{d}(\mathbf{l})$ has the property of *uniform shift* under first derivative (θ) change at θ_0; that is, show that

$$\left[\frac{\partial}{\partial\theta} g(l_1:\mathbf{d}, \theta)\right]_{\theta=\theta_0} = -\frac{\partial}{\partial l_1} g(l_1:\mathbf{d}, \theta_0).$$

(For details see Section 1.)

2. Let $F(x:\theta)$ be the Pareto distribution

$$F(x:\theta) = 1 - x^{-\theta}, \qquad x > 1, \quad \theta > 0.$$

Derive the linearized variable $l(x:\theta_0)$. Can the local linearization (at θ_0) be extended to a global linearization (all θ)? Extend if possible. (D. R. Brillinger, 1963.)

3. Let $F(x:\theta)$ be the Weibull distribution with density

$$f(x:\theta) = \frac{\beta_0}{\theta^{\beta_0}} x^{\beta_0 - 1} \exp\left\{-\left(\frac{x}{\theta}\right)^{\beta_0}\right\}, \qquad x > 0, \quad \theta > 0.$$

Derive the linearized variable $l(x:\theta_0)$. Can the local linearization be extended to a global linearization? Extend if possible. (J. Whitney.)

4. Let $F(x:\theta)$ be the chi-model with density

$$f(x:\theta) = \frac{1}{\Gamma(f/2)} \frac{x^{f-1}}{\theta^f 2^{f/2-1}} \exp\left\{-\frac{x^2}{2\theta^2}\right\}.$$

Derive the linearized variable $l(x:\theta_0)$. Can the local linearization be extended to global linearization? Extend if possible.

5. Let $\langle x \rangle$ be the nearest integer to x.

(i) If x has the continuous Poisson distribution $F(x:\theta)$ with quantity θ, check that $\langle x \rangle$ has the ordinary Poisson distribution with quantity θ. (Y. S. Lee.)

(ii) Check that

$$F(x + \tfrac{1}{2}:\theta) - F(x - \tfrac{1}{2}:\theta) = \frac{\theta^x}{\Gamma(x+1)} \exp\{-\theta\}$$

for all $x \geq 0$.

PART III

Extensions

CHAPTER SEVEN

Inference from Frequencies

The preceding chapters indicate the broad range of applications for the structural model. They also indicate some areas and directions that cannot be covered, or covered completely, by the structural model. For example, the additional quantity in Chapter Four cannot be described *directly* by the quantity of a structural model; the stochastically increasing model and the Poisson model in Chapter Six are not *exact* structural models.

Without any structuring relationship between the quantity and the response, there remains, in the discrete case, only the probability function to identify a response value and, in the continuous case, only the likelihood function to identify a response value (Section 1, Chapter Four).

The case of a discrete response variable is examined in this chapter. With multiple observations the composite response can be recorded alternatively by giving the *frequency* for each of the possible probability functions. For a large number of observations these frequencies acquire a position relationship to their probabilities in the same manner as with the Poisson model in Chapter Six. This provides the basis for large-sample structural inference.

The continuous response variable, with inference based on likelihood, is examined in Chapter Eight.

1 FREQUENCY MODELS AND THE POISSON BASIS

The *Poisson model* is perhaps the simplest statistical model with a frequency variable:

$$f(x:\theta) = \frac{\theta^x}{x!} \exp\{-\theta\}, \qquad x = 0, 1, 2, \ldots.$$

A slightly more complex model is the *binomial model*, which describes the frequencies x_1, x_2 of the occurrence, nonoccurrence of an event in n performances of a process:

$$f(x_1, x_2 : p_1, p_2) = \binom{n}{x_1 \quad x_2} p_1^{x_1} p_2^{x_2}, \qquad \begin{matrix} x_j \geq 0, \\ x_1 + x_2 = n, \end{matrix}$$

where $p_1 + p_2 = 1$ $(p_j \geq 0)$ and

$$\binom{n}{x_1 \ x_2} = \frac{n!}{x_1!\, x_2!} = \binom{n}{x_1} = \binom{n}{x_2}$$

is the combinatorial function.

A generalization is the *multinomial model*, which describes the frequencies $x_1, \ldots, x_r$ of occurrence of events $E_1, \ldots, E_r$ in n performances of a process, the events $E_1, \ldots, E_r$ for a performance being mutually exclusive and exhaustive:

$$f(\mathbf{x}:\mathbf{p}) = \binom{n}{x_1 \cdots x_r} p_1^{x_1} \cdots p_r^{x_r}, \qquad \begin{array}{c} x_j \geq 0, \\ \sum x_j = n, \end{array}$$

where $\sum p_j = 1$ $(p_j \geq 0)$ and

$$\binom{n}{x_1 \cdots x_r} = \frac{n!}{x_1! \cdots x_r!}$$

is the generalized combinatorial function. The quantity $\mathbf{p}$ may be restricted and may depend on an essential, but simpler, quantity θ: $\mathbf{p} = \mathbf{p}(\theta)$.

Several multinomial models can be combined to form a single *composite multinomial model:*

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_k : \mathbf{p}_1, \ldots, \mathbf{p}_k) = \prod_{i=1}^{k} \binom{n_i}{x_{1i} \cdots x_{r_i i}} p_{1i}^{x_{1i}} \cdots p_{r_i i}^{x_{r_i i}}, \qquad \begin{array}{c} x_{ji} \geq 0, \\ \sum_j x_{ji} = n_i, \end{array}$$

where $\sum_j p_{ji} = 1$ $(p_{ji} \geq 0)$ for each i. Again the quantities $\mathbf{p}_1, \ldots, \mathbf{p}_k$ may depend on a simpler quantity θ.

Now consider a finite population of N elements with m_1 elements of kind $E_1, \ldots, m_r$ elements of kind $E_r (\sum m_j = N)$. The *hypergeometric model* describes a succession of k random samples that exhaust the population; on the ith sample let $x_{1i}, \ldots, x_{ri}$ $(= \mathbf{x}_i)$ designate the frequencies of events $E_1, \ldots, E_r$; the ith sample size is n_i and $\sum n_i = N$:

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \frac{\prod n_i! \prod m_j!}{N! \prod x_{ji}!}, \qquad \begin{array}{c} x_{ji} \geq 0, \\ \sum_j x_{ji} = n_i, \\ \sum_i x_{ji} = m_j, \end{array}$$

$$= \frac{\prod_i \binom{n_i}{x_{1i} \cdots x_{ri}}}{\binom{N}{m_1 \cdots m_r}} = \frac{\prod_j \binom{m_j}{x_{j1} \cdots x_{jk}}}{\binom{N}{n_1 \cdots n_k}}.$$

The $m_1, \ldots, m_r$ may be quantities and may depend on an essential but simpler quantity θ.

The Poisson model provides a basis for analyzing the other models.

Consider r independent Poisson variables $x_1, \ldots, x_r$ with means $\varphi p_1, \ldots, \varphi p_r$ $(\sum p_j = 1)$. The composite model for the x's is

$$\frac{\prod (\varphi p_j)^{x_j}}{\prod x_j!} \exp\{-\varphi\};$$

the conditional model given that $\sum x_j = n$ is

$$\frac{\prod (\varphi p_j)^{x_j} \exp\{-\varphi\} / \prod x_j!}{\varphi^n \exp\{-\varphi\}/n!} = \binom{n}{x_1 \cdots x_r} p_1^{x_1} \cdots p_r^{x_r}.$$

The conditional model given $\sum x_j = n$ is the multinomial model in a preceding paragraph (see Figure 1). The multinomial model is obtained regardless of the value of φ. The choice, $\varphi = n$, is a simple and convenient choice: the vector of Poisson variables then has mean $(np_1, \ldots, np_r)$; the linear constraint $\sum x_j = n$ passes through the vector mean; and the vector mean of the Poisson variables is also the vector mean of the multinomial variable.

The composite multinomial is a combination of independent multinomials. The ith component multinomial can be obtained from Poisson variables with means $n_i p_{1i}, \ldots, n_i p_{r_i i}$ by imposing the condition $\sum_j x_{ji} = n_i$. The composite multinomial can then be obtained from k batches of Poisson variables $(i = 1, \ldots, k)$ by imposing the indicated constraint on each batch.

Now consider the k batches of Poisson variables but with $r_1 = \cdots = r_k = r$

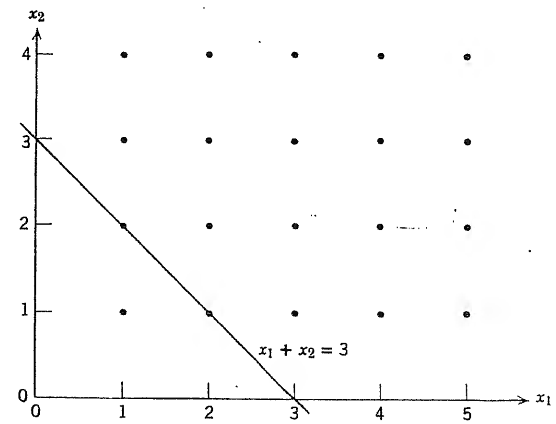


Figure 1 The possible values for two Poisson variables (x_1, x_2). The possible values for a binomial variable (x_1, x_2) with $n = 3$. The Poisson variables conditional on $x_1 + x_2 = 3$ give a binomial variable.

and with $\mathbf{p}_1 = \cdots = \mathbf{p}_k = \mathbf{p}$. The composite multinomial obtained by the conditions in the preceding paragraph can be further conditioned by $\Sigma_i x_{1i} = m_1, \ldots, \Sigma_i x_{ri} = m_r$. The resulting conditional model is the hypergeometric model described in a preceding paragraph. The hypergeometric does not depend on the vector $\mathbf{p}$; the vector $\mathbf{p}$ in the Poissons and multinomials can then be chosen so that the Poisson vector mean is also the mean of the hypergeometric ($\Sigma\, n_i p_1 = m_1, \ldots, \Sigma\, n_i p_r = m_r$).

The hypergeometric model can also be obtained from independent Poisson variables and an intermediate single multinomial. Let x_{ji} ($j = 1, \ldots, r$, $i = 1, \ldots, k$) be independent Poisson variables with means Np_{ji}, where $p_{ji} = m_j n_i / N^2$. A single multinomial is obtained from the condition $\Sigma\, x_{ji} = N$. The hypergeometric is obtained from the stronger conditions $\Sigma\, x_{j1} = n_1, \ldots, \Sigma\, x_{jk} = n_k$; $\Sigma\, x_{1i} = m_1, \ldots, \Sigma\, x_{ri} = m_r$.

2 FREQUENCY MODELS: LARGE SAMPLES

The common frequency models can be obtained by conditioning independent Poisson variables. In Chapter Six the analysis of the stochastically increasing model showed that the Poisson model, in a frequency application, was approximated by a simple measurement model, provided that the location quantity was large. This section describes how the Poisson result extends in a simple manner to cover the common frequency models, provided again that the location quantities are large.

Consider t Poisson variables $x_1, \ldots, x_t$ with means $Np_1, \ldots, Np_t$; and consider s linearly independent constraints with *integer* coefficients,

$$\sum l_{1i} x_i = c_1(N)$$
$$\vdots$$
$$\sum l_{si} x_i = c_s(N),$$

where the constants $c_j(N)$ are such that the mean vector $(Np_1, \ldots, Np_t)$ satisfies the constraints. The conditional distribution of the x's is a frequency model which for appropriate choice of constraints can be any of the models described in Section 1. This section shows that the conditional distribution of the x's, in a frequency application, can be approximated by a measurement model:

The error variables $e_1 = \sqrt{x_1} - \sqrt{Np_1}, \ldots, e_t = \sqrt{x_t} - \sqrt{Np_t}$ *have a distribution that approaches* (as $N \to \infty$) *the distribution of a sample of* t *from the normal with mean* 0 *and standard deviation* $\frac{1}{2}$ *but conditioned by the constraints as expressed in terms of the e's; the constraints approach linear form in any bounded range* (as $N \to \infty$).

A realized sequence $e_1, \ldots, e_t$ *from the error distribution provides the link between the measurements* $\sqrt{x_1}, \ldots, \sqrt{x_t}$ *and the location quantities* $\sqrt{Np_1}, \ldots, \sqrt{Np_t}$:

$$\sqrt{x_1} = \sqrt{Np_1} + e_1$$
$$\vdots$$
$$\sqrt{x_t} = \sqrt{Np_t} + e_t.$$

Let $e_1, \ldots, e_t$ be the error variable for the Poisson variables $x_1, \ldots, x_t$ (Section 2, Chapter Six):

$$e_1 = \sqrt{x_1} - \sqrt{Np_1}, \ldots, e_t = \sqrt{x_t} - \sqrt{Np_t};$$

and suppose $p_1 > 0, \ldots, p_t > 0$. By Section 3 in Chapter Six, the limiting distribution of the e's (as $N \to \infty$) is that of a sample of t from the normal with mean 0 and standard deviation $\frac{1}{2}$. As part of the derivation it was shown that the probability function for an error e (with a scale factor accommodating the average spacing between e values) approaches uniformly the density function for the normal with mean 0 and standard deviation $\frac{1}{2}$. From this it follows that the probability function for the error vector $\mathbf{e} = (e_1, \ldots, e_t)'$ (with a scale factor to accommodate the average spacing between $\mathbf{e}$ values) approaches uniformly the density for a sample of t from the normal with mean 0 and standard deviation $\frac{1}{2}$. It follows that, if attention is restricted to points $\mathbf{e}$ that satisfy the constraints, then the probability function for these points (with the same scale factor) approaches uniformly the density for the sample of t from the normal at these points. It then follows that the conditional distribution of the e's is as described by the measurement model—provided the following are established:

The constraints in terms of the e's become linear in any bounded range about (0, . . . , 0) (as $N \to \infty$) (see Figure 2).

The spacing of the $\mathbf{e}$ *points that satisfy the constraints becomes uniform in any bounded range, and the spacing between adjacent points goes to zero* (as $N \to \infty$).

The next paragraph proves the first of these statements: The proof is simple and the results are needed to summarize this section. The next section is devoted to proving the second statement: The proof is somewhat long and is not of general statistical interest.

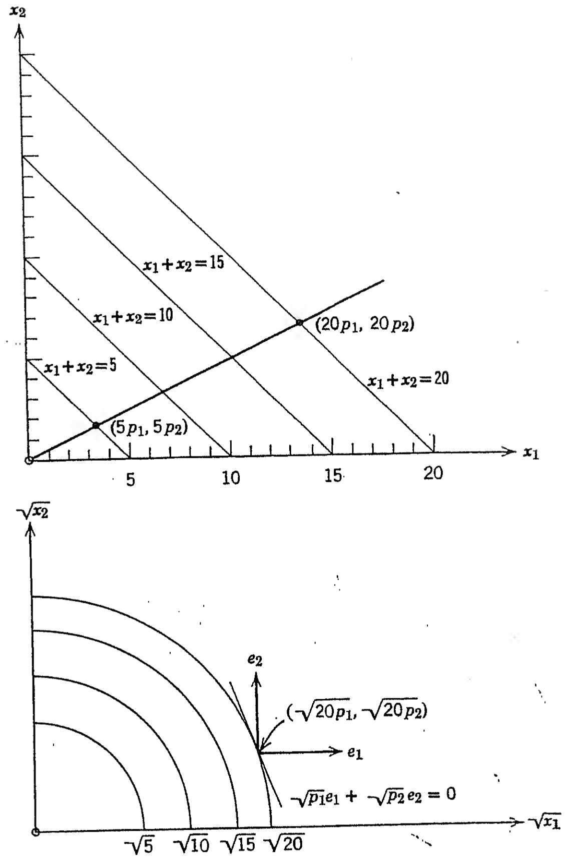


Figure 2 The binomial variable (x_1, x_2) with probabilities p_1, p_2, and $n = 5, 10, 15, 20$. The approximating error variable (e_1, e_2) with means 0, standard deviations $\frac{1}{2}$, and constraint $\sqrt{p_1}e_1 + \sqrt{p_2}e_2 = 0$.

Consider the form of the constraints,

$$\sum l_{1i}x_i = c_1(N)$$
$$\vdots$$
$$\sum l_{si}x_i = c_s(N),$$

in the neighborhood of the point $(Np_1, \ldots, Np_t)$ which satisfies the constraints. The error variables,

$$e_1 = \sqrt{x_1} - \sqrt{Np_1}, \ldots, e_t = \sqrt{x_t} - \sqrt{Np_t},$$

can be used to reexpress the constraints:

$$\sum l_{1i}(\sqrt{Np_i} + e_i)^2 = c_1(N)$$
$$\vdots$$
$$\sum l_{si}(\sqrt{Np_i} + e_i)^2 = c_s(N).$$

These can be simplified (the point $(Np_1, \ldots, Np_t)$ satisfies the constraints):

$$\sum l_{1i}\sqrt{p_i}\, e_i = -\sum l_{1i}e_i^2/2\sqrt{N}$$
$$\vdots$$
$$\sum l_{si}\sqrt{p_i}\, e_i = -\sum l_{si}e_i^2/2\sqrt{N}.$$

For any finite range for **e** these approach the linear constraints

$$\sum l_{1i}\sqrt{p_i}\, e_i = 0$$
$$\vdots$$
$$\sum l_{si}\sqrt{p_i}\, e_i = 0$$

as $N \to \infty$ (see Figure 2).

Consider briefly the quantity **p**, and suppose that it depends in a continuously differentiable way on a simpler quantity $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$. Let $\boldsymbol{\theta}^0$ be a reference value for $\boldsymbol{\theta}$ and consider the model for $\boldsymbol{\theta}$ near $\boldsymbol{\theta}^0$. The square root of $p_i(\boldsymbol{\theta})$ can be expanded by Taylor's theorem:

$$\sqrt{p_i(\boldsymbol{\theta})} = \sqrt{p_i(\boldsymbol{\theta}^0)} + \sum_{u=1}^{r}(\theta_u - \theta_u^0)\frac{1}{2\sqrt{p_i(\boldsymbol{\theta}^0)}}\frac{\partial p_i(\boldsymbol{\theta}^0)}{\partial \theta_u} + \cdots.$$

Hence

$$(\sqrt{Np_1(\boldsymbol{\theta})}, \ldots, \sqrt{Np_t(\boldsymbol{\theta})})$$
$$= (\sqrt{Np_1(\boldsymbol{\theta}^0)}, \ldots, \sqrt{Np_t(\boldsymbol{\theta}^0)}) + (\theta_1 - \theta_1^0)\mathbf{v}_1' + \cdots + (\theta_r - \theta_r^0)\mathbf{v}_r' + \cdots,$$

where†

$$\mathbf{v}_1 = \left(\frac{\partial\sqrt{Np_1(\boldsymbol{\theta}^0)}}{\partial\theta_1}, \ldots, \frac{\partial\sqrt{Np_t(\boldsymbol{\theta}^0)}}{\partial\theta_1}\right)'$$

$$\vdots$$

$$\mathbf{v}_r = \left(\frac{\partial\sqrt{Np_1(\boldsymbol{\theta}^0)}}{\partial\theta_r}, \ldots, \frac{\partial\sqrt{Np_t(\boldsymbol{\theta}^0)}}{\partial\theta_r}\right)'.$$

If the location quantity

$$(\sqrt{Np_1(\boldsymbol{\theta})}, \ldots, \sqrt{Np_t(\boldsymbol{\theta})})$$

is within two or three standard deviations of the reference value

$$(\sqrt{Np_1(\boldsymbol{\theta}^0)}, \ldots, \sqrt{Np_t(\boldsymbol{\theta}^0)}),$$

then the corresponding quantity $\boldsymbol{\theta}$ differs from $\boldsymbol{\theta}^0$ by an amount of order $N^{-1/2}$. It follows that a first derivative approximation is appropriate as N becomes large. It follows also that the deviation of the location quantity

$$(\sqrt{Np_1(\boldsymbol{\theta})}, \ldots, \sqrt{Np_t(\boldsymbol{\theta})})$$

from the reference value

$$(\sqrt{Np_1(\boldsymbol{\theta}^0)}, \ldots, \sqrt{Np_t(\boldsymbol{\theta}^0)})$$

is approximately linear in terms of structural vectors $\mathbf{v}_1', \ldots, \mathbf{v}_r'$ with coefficients $(\theta_1 - \theta_1^0), \ldots, (\theta_r - \theta_r^0)$.

The results in this section can now be summarized. Let $x_1, \ldots, x_t$ be frequency variables that satisfy the linearly independent constraints

$$\sum l_{1i}x_i = c_1(N)$$

$$\vdots$$

$$\sum l_{si}x_i = c_s(N)$$

with integer coefficients. And suppose the model for the frequency variables is that of independent Poisson variables $x_1, \ldots, x_t$ that satisfy the constraints and have means $Np_1(\boldsymbol{\theta}), \ldots, Np_t(\boldsymbol{\theta})$ that satisfy the constraints ($p_j > 0$ and $\mathbf{p}$ continuously differentiable near $\boldsymbol{\theta}^0$). Then, for N large, the frequency model

† Notation for the derivative at a particular point: $df(\theta^0)/d\theta = [df(\theta)/d\theta]_{\theta=\theta^0}$.

with $\boldsymbol{\theta}$ near $\boldsymbol{\theta}^0$ can be approximated by the following *measurement model, the simple regression model* (Problem 17, Chapter Three):

$$e_1, \ldots, e_t,$$

$$\sqrt{x_1} - \sqrt{Np_1(\boldsymbol{\theta}^0)} = \sum_{u=1}^{r}(\theta_u - \theta_u^0)v_{u1} + e_1$$

$$\vdots$$

$$\sqrt{x_t} - \sqrt{Np_t(\boldsymbol{\theta}^0)} = \sum_{u=1}^{r}(\theta_u - \theta_u^0)v_{ut} + e_t.$$

The model has an error vector $(e_1, \ldots, e_t)$ *which has the distribution of a sample from the normal with mean* 0 *and standard deviation* $\frac{1}{2}$ *but conditioned to satisfy the constraints*

$$\sum l_{1i}\sqrt{p_i(\boldsymbol{\theta}^0)} \cdot e_i = 0$$

$$\vdots$$

$$\sum l_{si}\sqrt{p_i(\boldsymbol{\theta}^0)} \cdot e_i = 0,$$

and the model has a structural equation in which a realized error vector provides the link between the measurement deviation from the reference location and the quantity deviations $\theta_1 - \theta_1^0, \ldots, \theta_r - \theta_r^0$ *(as coefficients of the structural vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_r$*).*

This form of measurement model can be analyzed in a straightforward manner by the methods of Chapter Three as applied in Problems 17, 18 in that chapter. Examples are given in Sections 4, 5, 6.

The measurement model is an approximate model, a limiting model as $N \to \infty$. In an application, with a given N, the curvature of the constraints and the curvature of the mean vector $\mathbf{p}(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$ may not be negligible. Analysis in the neighborhood of a reference value $\boldsymbol{\theta}^0$ may indicate the quantity $\boldsymbol{\theta}$ to be near a value $\boldsymbol{\theta}^{(1)}$ beyond the range of reasonable linearity. The value $\boldsymbol{\theta}^{(1)}$ can then be used as a new reference value. Analysis in the neighborhood of $\boldsymbol{\theta}^{(1)}$ may then indicate the quantity $\boldsymbol{\theta}$ to be near $\boldsymbol{\theta}^{(2)}$. The procedure can be repeated, forming steps of an iteration. After several steps, in the typical application, the approximating model will describe probabilities for $\boldsymbol{\theta}$ in the linear neighborhood of the final reference point.

*3 THE UNIFORMITY PROOF

This section completes the proof in the preceding section by establishing the uniformity of points that satisfy certain linear constraints. The proof is

not of central statistical interest and may be omitted without affecting the background for succeeding sections.

The range of the Poisson variable $\mathbf{x} = (x_1, \ldots, x_t)'$ is the set of positive *integer* or *lattice* points in R^t:

$$S^+ = \{\mathbf{x}: \quad x_i \geq 0, x_i = \text{integer}, i = 1, \ldots, t\}.$$

For the earlier parts of the proof it is more convenient to work with the set of all lattice points

$$S = \{\mathbf{x}: \quad x_i = \text{integer}, i = 1, \ldots, t\}.$$

The points of the lattice can be represented in terms of an Abelian group of integer translations. Let

$$g_1\mathbf{x} = (x_1 + 1, x_2, \ldots, x_t)'$$
$$\vdots$$
$$g_t\mathbf{x} = (x_1, \ldots, x_{t-1}, x_t + 1)';$$

then

$$g_1^{y_1} \cdots g_t^{y_t}\mathbf{x} = (x_1 + y_1, \ldots, x_t + y_t)'$$

(each y is an integer). It follows that the set

$$G = \{g_1^{y_1} \cdots g_t^{y_t}: \quad y_i = \text{integer } (i = 1, \ldots, t)\}$$

is an Abelian group. The group can be used to provide coordinates on S: Let $\mathbf{0} = (0, \ldots, 0)'$ be the reference point; let

$$[\mathbf{x}] = g_1^{x_1} \cdots g_t^{x_t};$$

then the point $\mathbf{x}$ has position $[\mathbf{x}]$ relative to the reference point $\mathbf{0}$,

$$\mathbf{x} = [\mathbf{x}]\mathbf{0}.$$

The lattice points of S are uniformly spaced in R^n. The transformations of G express this uniformity: A transformation $g_1^{y_1} \cdots g_t^{y_t}$ maps S onto itself by a translation of y_1 units in the direction of the first axis, $\ldots$, y_t units in the direction of the tth axis. The group demonstrates that the set S has the same form at all points: *The set S is homogeneous relative to the translation group G.*

The linearly independent constraints

$$\sum l_{1i}x_i = c_1(N)$$
$$\vdots$$
$$\sum l_{si}x_i = c_s(N)$$

can be examined for various values of the c's; and they can then be viewed as providing the first s of an alternative set of coordinates for R^t:

$$w_1 = l_{11}x_1 + \cdots + l_{1t}x_t$$
$$\vdots$$
$$w_s = l_{s1}x_1 + \cdots + l_{st}x_t$$
$$\vdots$$
$$w_t = l_{t1}x_1 + \cdots + l_{tt}x_t;$$

the remaining $t - s$ coordinates can be based on a completing set of $t - s$ linearly independent constraints with integer coefficients. A lattice point $\mathbf{x}$ (an $\mathbf{x}$ vector in S) can be represented in alternative coordinates as a $\mathbf{w}$ vector (the corresponding $\mathbf{w}$ vector has integer coefficients). For an arbitrary $\mathbf{x}$ vector in R^t there is a corresponding $\mathbf{w}$ vector. An arbitrary $\mathbf{x}$ vector has a neighboring lattice point; and correspondingly an arbitrary $\mathbf{w}$ vector has a neighboring lattice point (see Figure 3).

Now consider the lattice points that satisfy the s constraints with right sides set equal to zero:

$$S_C = \left\{\mathbf{x}: \quad \sum_i l_{ji}x_i = 0, \quad j = 1, \ldots, s\right\}.$$

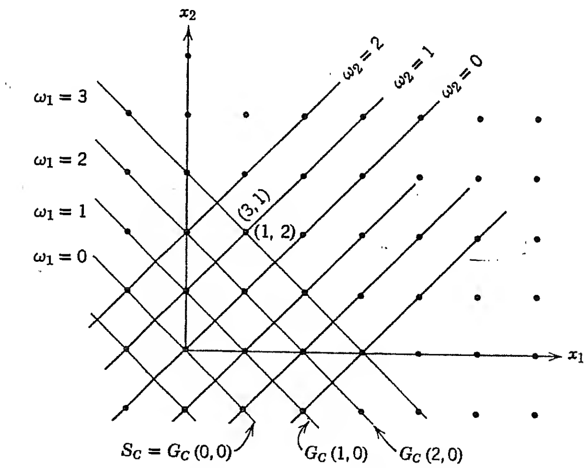


Figure 3 New coordinates $w_1 = x_1 + x_2$, $w_2 = x_2 - x_1$. A point in old coordinates (1, 2); the point in new coordinates (3, 1). The set S_C based on the condition $x_1 + x_2 = 0$: S_C is the orbit of a subgroup $G_C = \{g_1^y g_2^{-y} : y = \text{integer}\}$.

In terms of the group coordinates this is the set

$$G_C = \left\{ g_1^{y_1} \cdots g_t^{y_t}: \ \sum_i l_{ji} y_i = 0, \quad j = 1, \ldots, s \right\}.$$

But the set G_C is closed under the formation of products and inverses; it is a subgroup. A subgroup in an Abelian group is a normal subgroup.

The subgroup G_C partitions the set S into orbits $G_C\mathbf{x}$ ($\mathbf{x}$ in S). The orbit through the origin $\mathbf{0}$ is

$$G_C\mathbf{0} = \left\{ \mathbf{x}: \ \sum_i l_{ji} x_i = 0, \quad \mathbf{x} \in S, \quad j = 1, \ldots, s \right\};$$

the orbit through a point $\mathbf{x}^0$ is

$$G_C\mathbf{x}^0 = \left\{ \mathbf{x}: \ \sum_i l_{ji} x_i = \sum_i l_{ji} x_i^0, \quad \mathbf{x} \in S, \quad j = 1, \ldots, s \right\}.$$

Any translation in G_C carries an orbit into itself: The orbit is homogeneous under the group G_C; the orbit has the same form at all points. The group G_C generates points spread uniformly in the $(t - s)$-dimensional subspace:

$$S_C^* = \left\{ \mathbf{x}: \ \sum_i l_{ji} x_i = 0, \quad \mathbf{x} \in R^t, \quad j = 1, \ldots, s \right\};$$

and by translation the group G_C generates points spread uniformly in the $(t - s)$-dimensional subset:

$$\left\{ \mathbf{x}: \ \sum_i l_{ji} x_i = \sum_i l_{ji} x_i^0, \quad \mathbf{x} \in R^t, \quad j = 1, \ldots, s \right\}$$

passing through a point $\mathbf{x}^0$ of S.

Now consider how the uniformity of lattice points $\mathbf{x}$ in a subset satisfying the constraints carries over to uniformity of the corresponding $\mathbf{e}$ points in a bounded neighbourhood of $\mathbf{0}$. The volume change from a point $\mathbf{e}$ to an original point $\mathbf{x}$ is

$$\begin{aligned} \prod_{i=1}^{t} 2(\sqrt{Np_i} + e_i) &= 2^t N^{t/2} \prod_i (\sqrt{p_i} + e_i/\sqrt{N}) \\ &= 2^t N^{t/2} h_N(\mathbf{e}), \end{aligned}$$

where $h_N(\mathbf{e})$ approaches $\prod \sqrt{p_i}$ uniformly in any bounded range as $N \to \infty$. Thus the uniformity of the constrained lattice points in a region surrounding

$(Np_1, \ldots, Np_t)$ becomes uniformity for the points $\mathbf{e}$ about $\mathbf{0}$; and the spacing between $\mathbf{e}$ points goes to zero as $N \to \infty$.

This establishes the uniformity property required in the preceding section.

4 THE MULTINOMIAL MODEL

Consider an example involving a multinomial model. The theory in Section 2 relates the multinomial model to a measurement model with normal errors and with structural vectors; the methods of Chapter Three can then be applied.

Consider two factors affecting the breeding of maize: a first factor, *Starchy* S or *sugary* s; a second factor, *Green* G or *white* g. The data record the classification of $n = 3839$ progeny of self-fertilized heterozygotes:

	G	g	Total
S	1997	906	2903
s	904	32	936
Total	2901	938	3839.

The accepted theory for the example prescribes marginal probabilities in the ratio 3:1 for S to s and in the ratio 3:1 for G to g; but it allows for a genetic factor *linkage* involving a quantity θ so that cell probabilities can differ from the independence pattern,

	G	g	
S	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{4}$
s	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
	$\frac{3}{4}$	$\frac{1}{4}$	1,

and have the form

	G	g	
S	$\frac{1}{4}(2 + \theta)$	$\frac{1}{4}(1 - \theta)$	$\frac{3}{4}$
s	$\frac{1}{4}(1 - \theta)$	$\frac{1}{4}\theta$	$\frac{1}{4}$
	$\frac{3}{4}$	$\frac{1}{4}$	1.

The linkage quantity θ can have any value in the interval (0, 1); the value $\theta = \frac{1}{4}$ corresponds to independence (no linkage). This suggests analyzing the model with the linkage quantity, but allowing for the possibility of withdrawal to the simpler model corresponding to independence.

The data can be transformed to *measurements and location quantities:*

	Observed	$\sqrt{\text{Observed}}$	$\sqrt{\text{Mean}}$	$\frac{d}{d\theta}\sqrt{\text{Mean}}$
	x	$\mathbf{y}$	$\boldsymbol{\tau}(\theta)$	$\mathbf{v}(\theta)$
SG	1997	44.687,806	$\sqrt{n}\,\frac{1}{2}\sqrt{2+\theta}$	$\sqrt{n}/4\sqrt{2+\theta}$
Sg	906	30.099,834	$\sqrt{n}\,\frac{1}{2}\sqrt{1-\theta}$	$-\sqrt{n}/4\sqrt{1-\theta}$
sG	904	30.066,593	$\sqrt{n}\,\frac{1}{2}\sqrt{1-\theta}$	$-\sqrt{n}/4\sqrt{1-\theta}$
sg	32	5.656,854	$\sqrt{n}\,\frac{1}{2}\sqrt{\theta}$	$\sqrt{n}/4\sqrt{\theta}$
	3839	61.959,664		

For a reference value $\theta^0 = 0.1$ the measurement vector and the structural vector are

$\mathbf{y}$	$\mathbf{v}(0.1)$
44.687,806	10.689,058
30.099,834	−16.327,805
30.066,593	−16.327,805
5.656,854	48.983,416.

The structural equation for the measurement model at $\theta^0 = 0.1$ is

$$\mathbf{y} - \boldsymbol{\tau}(0.1) = (\theta - 0.1)\mathbf{v}(0.1) + \mathbf{e}.$$

The appropriate regression coefficient is

$$b^{(1)} = \frac{(\mathbf{y} - \boldsymbol{\tau}(0.1), \mathbf{v}(0.1))}{(\mathbf{v}(0.1), \mathbf{v}(0.1))} = \frac{(\mathbf{y}, \mathbf{v}(0.1))}{(\mathbf{v}(0.1), \mathbf{v}(0.1))}$$

$$= \frac{-227.623,11}{3046.825,3} = -0.074,708$$

(the simplification from the measurement deviation $\mathbf{y} - \boldsymbol{\tau}(0.1)$ to measurement $\mathbf{y}$ is based on the orthogonality property:

$$\sum \tau_i^2(\theta) = n,$$

$$\sum 2\tau_i(\theta)\frac{d}{d\theta}\tau_i(\theta) = 0,$$

$$(\boldsymbol{\tau}(\theta), \mathbf{v}(\theta)) = 0;$$

the same kind of simplification will occur throughout the examples). The corresponding θ value based on linearity at the reference point is then

$$\theta^{(1)} = 0.1 + b^{(1)} = 0.025,292.$$

The value $\theta^{(1)} = 0.025,292$ can be used as a new reference value. The measurement vector $\mathbf{y}$ and the structural vector $\mathbf{v}(\theta^{(1)})$ are

$\mathbf{y}$	$\mathbf{v}$
44.687,806	10.884,419
30.099,834	−15.689,595
30.066,593	−15.689,595
5.656,854	97.400,229;

the corresponding regression coefficient and value of the quantity are

$$b^{(2)} = \frac{(\mathbf{y}, \mathbf{v})}{(\mathbf{v}, \mathbf{v})} = \frac{93.392,834}{10,097.601} = 0.009,249,$$

$$\theta^{(2)} = 0.034,541.$$

The nonlinearity is prominent in the fourth coordinate with mean $\frac{1}{2}\sqrt{n}\sqrt{\theta}$ and with θ near zero. Three further iterations effectively overcome this nonlinearity:

i	$b^{(i)}$	$\theta^{(i)}$
0		0.100,000
1	−0.074,708	0.025,292
2	0.009,249	0.034,541
3	0.001,094	0.035,635
4	0.000,042	0.035,677
5	0.000,001	0.035,678.

The variance of the regression coefficient for error can be obtained in part from the inverse "matrix" for the last iteration:

	$\mathbf{v}$	$\mathbf{y}$	
$\mathbf{v}$	7348.777,2	0.009,168	1
$\mathbf{y}$	1	0.000,001	0.000,136,077.

The basic error variance is $\frac{1}{4}$; the error variables satisfy one linear constraint; the normal error distribution is rotationally symmetric; the error variance is therefore $\frac{1}{4}$ for three orthonormal variables in the subspace satisfying the constraint. It follows that the variance of the regression coefficient for the error variable is

$$\tfrac{1}{4}(0.000,136) = 0.000,034,$$

and its standard deviation is 0.005,83.

The model with no linkage ($\theta = \frac{1}{4}$) has no quantity; the location values are obtained by substituting $\theta = \frac{1}{4}$ in the expression for $\boldsymbol{\tau}(\theta)$.

The observed and fitted vectors (and the relevant difference vectors) can be exhibited in a table:

Measurement $\mathbf{y}$		Fitted Location (Allowing Linkage with $\theta = 0.035,678$) $\boldsymbol{\tau}$		Fitted Location (No Linkage $\theta = 0.25$) $\boldsymbol{\tau}$
44.687,806	0.486,652	44.201,154	−2.268,594	46.469,748
30.099,834	−0.322,330	30.422,164	3.592,843	26.829,321
30.066,593	−0.355,571	30.422,164	3.592,843	26.829,321
5.656,854	−0.194,809	5.851,663	−9.638,253	15.489,916
	$l^2 = 0.505,108$		$l^2 = 123.859,481$	

The table records the squared lengths of the difference vectors.

The results can be summarized in an analysis-of-variance table:

Source	Dimension	Component	χ^2
Linkage	1	123.859,481	495.438
Deviations	2	0.505,108	2.020
Error ($\sigma_0^2 = \frac{1}{4}$)			
Total	3		

The error variance is $\sigma_0^2 = \frac{1}{4}$. The components can accordingly be adjusted to give chi-square values:

$$\chi^2 = \frac{\text{Component}}{\frac{1}{4}} = 4 \text{ Component}.$$

The observed chi-square value 2.020 falls between the 40% point (1.83) and the 30% point (2.41) for a chi-square variable on two degrees of freedom. The observed value 2.020 is thus a reasonable value for such a variable, and it indicates that the data are in accord with the linkage model.

The chi-square value 495.438 is an *extreme* value: For a chi-square variable on one degree of freedom the 1% point is 6.635 and the 0.1% point is 10.827. Thus within the linkage model there is strong evidence that θ is different from the no-linkage value $\frac{1}{4}$.

The structural distribution for the quantity θ is normal with mean 0.035,678 and standard deviation 0.005,83.

Consider briefly the special multinomial model, *the binomial model*. For the binomial model the frequency variables are x_1 and $n - x_1 = x_2$ and the probabilities are p and $1 - p = q$. The corresponding measurement variables and location quantities are

$$y_1 = \sqrt{x_1}, \qquad \tau_1 = \sqrt{np},$$
$$y_2 = \sqrt{n - x_1} = \sqrt{x_2}, \qquad \tau_2 = \sqrt{n - np} = \sqrt{nq}.$$

The approximating measurement model is

$$e_1, e_2,$$
$$y_1 = \tau_1 + e_1,$$
$$y_2 = \tau_2 + e_2;$$

the error variables e_1, e_2 are normal with mean 0, with standard deviation $\frac{1}{2}$ and subject to the constraint

$$\sqrt{p^0}\, e_1 + \sqrt{1 - p^0}\, e_2 = 0$$

in the neighborhood of p^0, $1 - p^0$ (see Figure 4).

An alternative approximating model can be formed by having a single

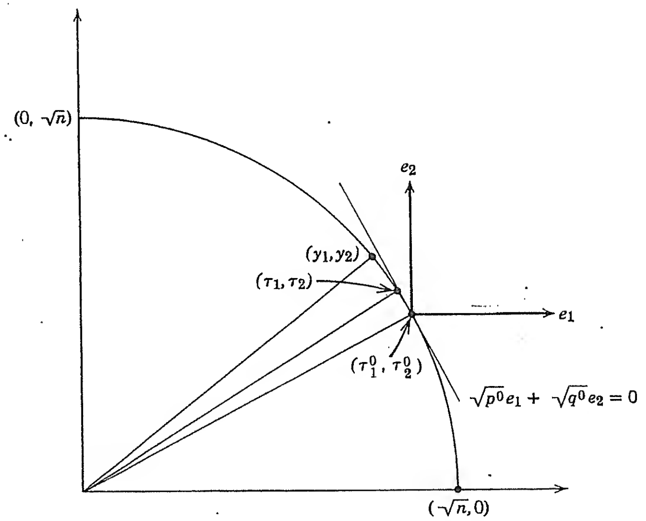


Figure 4 The reference point $(\tau_1^0, \tau_1^0) = (\sqrt{np^0}, \sqrt{nq^0})$; the location quantity $(\tau_1, \tau_2) = (\sqrt{np}, \sqrt{nq})$; and the measurement $(y_1, y_2) = (\sqrt{x_1}, \sqrt{n - x_1})$.

error variable and no constraint. The position of points on the quarter circle can be described by distance along the arc commencing at $(0, \sqrt{n})$:

$$\text{measurement} = \sqrt{n}\,\sin^{-1}\sqrt{\frac{x_1}{n}},$$

$$\text{location quantity} = \sqrt{n}\,\sin^{-1}\sqrt{p}.$$

Let e be a normal error variable with mean 0 and standard deviation $\frac{1}{2}$. The alternative model is

$$e$$

$$\sqrt{n}\,\sin^{-1}\sqrt{\frac{x_1}{n}} = \sqrt{n}\,\sin^{-1}\sqrt{p} + e.$$

The structural equation can conveniently be used in the form

$$\sin^{-1}\sqrt{\frac{x_1}{n}} = \sin^{-1}\sqrt{p} + \frac{e}{\sqrt{n}}.$$

The transformation $\sin^{-1}\sqrt{t}$ is recorded in most collections of statistical tables.

5 THE COMPOSITE MULTINOMIAL MODEL

Several independent multinomial models can be combined to form a composite multinomial model. The methods of analysis for the multinomial model in the preceding section indicate the pattern for the composite multinomial.

For example, consider some data on blood types. A sample of 353 people from a community C are classified by blood phenotype: O, A, B, AB; and a sample of 364 people from a second community D are similarly classified:

	C	D	Total
O	121	118	239
A	120	95	215
B	79	121	200
AB	33	30	63
Total	353	364	717.

The observable *phenotype* corresponds to an unobservable or latent *genotype*. An O gene is recessive to an A gene and to a B gene. Let p, q, r be the gene probabilities corresponding to A, B, O ($p + q + r = 1$). With random

mating the genotype and phenotype probabilities are

Genotype		Phenotype	
OO	r^2	O	r^2
AA	p^2	A	$p(p + 2r)$
AO	$2pr$		
BB	q^2	B	$q(q + 2r)$
BO	$2qr$		
AB	$2pq$	AB	$2pq$
	1		1

The model contains effectively two quantities p, q ($r = 1 - p - q$).

The phenotype model can be analyzed by allowing different gene probabilities for the two communities (in effect, four quantities). Withdrawal can then be made to the more specialized model having the same gene probabilities in the two communities. A further withdrawal might then be made to a model with specified values for the gene probabilities.

Consider first the phenotype model for community C:

	Observed x	$\sqrt{\text{Observed}}$ y	$\sqrt{\text{Mean}}$ τ	$\frac{\partial\tau}{\partial p}$ v_1	$\frac{\partial\tau}{\partial q}$ v_2
O	121	11.000,000	$\sqrt{nr^2}$	$-r\frac{353}{\tau_O}$	$-r\frac{353}{\tau_O}$
A	120	10.954,451	$\sqrt{np(p+2r)}$	$r\frac{353}{\tau_A}$	$-p\frac{353}{\tau_A}$
B	79	8.888,194	$\sqrt{nq(q+2r)}$	$-q\frac{353}{\tau_B}$	$r\frac{353}{\tau_B}$
AB	33	5.744,563	$\sqrt{n2pq}$	$q\frac{353}{\tau_{AB}}$	$p\frac{353}{\tau_{AB}}$
	353	18.788,294			

The derivatives are straightforward:

$$\frac{\partial r^2}{\partial p} = 2r(-1),$$

$$\frac{\partial(p^2 + 2pr)}{\partial p} = \frac{\partial}{\partial p}[p^2 + 2p(1 - p - q)] = 2r.$$

As a reference point consider $(p^0, q^0) = (\frac{1}{3}, \frac{1}{3})$; the structural vectors and corresponding regression coefficients for a *first* calculation, based on this

reference point, are:

y	v_1	v_2
11.000,000	−18.788,294	−18.788,294
10.954,451	10.847,427	−10.847,427
8.888,194	−10.847,427	10.847,427
5.744,563	13.285,330	13.285,330

$$(v_1, v_1)b_1 + (v_1, v_2)b_2 = (v_1, y),$$
$$(v_2, v_1)b_1 + (v_2, v_2)b_2 = (v_2, y),$$
$$b_1^{(1)} = -0.075{,}469, \qquad b_2^{(1)} = -0.170{,}711.$$

The corresponding values for p, q based on linearity at $p^0 = \frac{1}{3}$, $q^0 = \frac{1}{3}$ are

$$\begin{aligned} p^{(1)} &= 0.333{,}333 & q^{(1)} &= 0.333{,}333 \\ &\;\;\,-0.075{,}469 & &\;\;\,-0.170{,}711 \\ &= 0.257{,}864 & &= 0.162{,}622 \end{aligned}$$

The point $(p^{(1)}, q^{(1)})$ can be used as a reference point for a second calculation.

Successive iterations provide a check for nonlinearity. The following table records successive reference points (differences in parentheses):

i	$p^{(i)}$	$q^{(i)}$
0	0.333,333	0.333,333
	(−0.075,469)	(−0.170,711)
1	0.257,864	0.162,622
	(−0.011,691)	(0.010,465)
2	0.246,173	0.173,087
	(0.000,264)	(−0.000,004)
3	0.246,437	0.173,083
	(0.000,000)	(0.000,003)
4	0.246,437	0.173,086.

The variances and covariance of the error regression coefficients are equal to the appropriate elements of the inverse matrix multiplied by $\frac{1}{4}$. The inverse matrix at the last iteration is

$$\begin{pmatrix} (v_1, v_1) & (v_1, v_2) \\ (v_2, v_1) & (v_2, v_2) \end{pmatrix}^{-1} = \begin{pmatrix} 865.709{,}17 & 230.305{,}04 \\ 230.305{,}04 & 1{,}181.253{,}27 \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} 0.001{,}218 & -0.000{,}238 \\ -0.000{,}238 & 0.000{,}893 \end{pmatrix}.$$

The variances for the error regression coefficients for p, q are 0.000,304, 0.000,223; the standard deviations are 0.0174, 0.0149; and the correlation is −0.23.

The analysis for community D proceeds similarly:

	Observed x	$\sqrt{\text{Observed}}$ y	$\sqrt{\text{Mean}}$ τ	$\frac{\partial\tau}{\partial p}$ v_1	$\frac{\partial\tau}{\partial q}$ v_2
O	118	10.862,780	$\sqrt{nr^2}$	$-r\frac{364}{\tau_O}$	$-r\frac{364}{\tau_O}$
A	95	9.746,794	$\sqrt{np(p+2r)}$	$r\frac{364}{\tau_A}$	$-p\frac{364}{\tau_A}$
B	121	11.000,000	$\sqrt{nq(q+2r)}$	$-q\frac{364}{\tau_B}$	$r\frac{364}{\tau_B}$
AB	30	5.477,226	$\sqrt{n2pq}$	$q\frac{364}{\tau_{AB}}$	$p\frac{364}{\tau_{AB}}$
	364	19.078,784			

As an initial reference point consider $(p, q) = (0.2, 0.2)$, a point suggested by the final reference point for community C:

y	v_1	v_2
10.862,780	−19.078,784	−19.078,784
9.746,794	21.633,308	−7.211,103
11.000,000	−7.211,103	21.633,308
5.477,226	13.490,738	13.490,738

$$b_1^{(1)} = -0.009{,}223 \qquad b_2^{(1)} = 0.034{,}224$$
$$p^{(1)} = 0.190{,}777 \qquad q^{(1)} = 0.234{,}224$$

Successive iterations provide a check for nonlinearity:

i	$p^{(i)}$	$q^{(i)}$
0	0.200,000	0.200,000
	(−0.009,223)	(0.034,224)
1	0.190,777	0.234,224
	(−0.000,565)	(0.001,453)
2	0.190,212	0.235,677
	(−0.000,017)	(0.000,011)
3	0.190,195	0.235,688

The variances and covariance for the error regression coefficients are equal to the appropriate elements of the inverse matrix multiplied by $\frac{1}{4}$. The inverse matrix at the last iteration is

$$\begin{bmatrix} 1122.751{,}62 & 238.859{,}21 \\ 238.859{,}21 & 930.472{,}18 \end{bmatrix}^{-1} = \begin{bmatrix} 0.000{,}942 & -0.000{,}242 \\ -0.000{,}242 & 0.001{,}137 \end{bmatrix}.$$

The variances for the error regression coefficients for p, q are 0.000,236, 0.000,284; the standard deviations are 0.0154, 0.0169; and the correlation is -0.23.

Now consider the model based on having the same gene probabilities for the two communities. Points having the same probability function should be combined; accordingly the data for the two communities are combined:

	Observed	$\sqrt{\text{Observed}}$	$\sqrt{\text{Mean}}$	$\dfrac{\partial\tau}{\partial p}$	$\dfrac{\partial\tau}{\partial q}$
	$\mathbf{x}$	$\mathbf{y}$	$\boldsymbol{\tau}$	$\mathbf{v}_1$	$\mathbf{v}_2$
O	239	15.459,625	$\sqrt{nr^2}$	$-r\dfrac{717}{\tau_O}$	$-r\dfrac{717}{\tau_O}$
A	215	14.662,878	$\sqrt{np(p+2r)}$	$r\dfrac{717}{\tau_A}$	$-p\dfrac{717}{\tau_A}$
B	200	14.142,136	$\sqrt{nq(q+2r)}$	$-q\dfrac{717}{\tau_B}$	$r\dfrac{717}{\tau_B}$
AB	63	7.937,254	$\sqrt{n2pq}$	$q\dfrac{717}{\tau_{AB}}$	$p\dfrac{717}{\tau_{AB}}$
	717	26.776,856			

As an initial reference point consider $(p, q) = (0.22, 0.20)$; from the reference value $p = 0.22$, $q = 0.20$, $r = 0.58$, successive iterations give

i	$p^{(i)}$	$q^{(i)}$
0	0.220,000	0.200,000
	$(-0.002{,}379)$	$(0.004{,}327)$
1	0.217,621	0.204,327
	$(-0.000{,}007)$	$(0.000{,}025)$
2	0.217,614	0.204,352

The observed and fitted values can be compared by means of vectors in R^8:

	Measurement $\mathbf{y}$		Fitted Location $\boldsymbol{\tau}$		Fitted Location $\boldsymbol{\tau}$
Community C ($n = 352$)			$p = 0.246{,}437$ $q = 0.173{,}086$		$p = 0.217{,}614$ $q = 0.204{,}352$
O	11.0000	0.0938	10.9062	0.0459	10.8603
A	10.9545	-0.1104	11.0649	0.7925	10.2724
B	8.8882	-0.1400	9.0282	-0.8781	9.9063
AB	5.7446	0.2570	5.4876	-0.1156	5.6032
Community D ($n = 364$)			$p = 0.190{,}195$ $q = 0.235{,}688$		$p = 0.217{,}614$ $q = 0.204{,}352$
O	10.8628	-0.0907	10.9535	-0.0747	11.0282
A	9.7468	0.1208	9.6260	-0.8053	10.4313
B	11.0000	0.1038	10.8962	0.8367	10.0595
AB	5.4772	-0.2354	5.7126	0.0228	5.6898
		$l_C^2 = 0.1066$ $l_D^2 = 0.0890$		$l_{C,D}^2 = 2.7693$	

The fitted vector for the combined communities uses $n = 352$ for the first four coordinates (community C) and $n = 364$ for the last four coordinates (community D). The difference vectors are recorded together with squared lengths. The analysis-of-variance table is

Source	Dimension	Component	χ^2
Between communities	2	2.7693	11.077
Deviations for community C	1	0.1066	0.426
Deviations for community D	1	0.0890	0.356
Error ($\sigma_0^2 = \frac{1}{4}$)			
Total	4		

The chi-square values 0.426 and 0.356 are close to the 50% value (0.455) for a chi-square variable on one degree of freedom. These values are reasonable values; they indicate *for each community* that the data are in accord with the model.

The chi-square value 11.077 is beyond the 1% value (9.21) for a chi-square variable on two degrees of freedom. This is moderately strong evidence that the gene frequencies are different in the two communities.

The structural distribution for p and q in community C is normal:

	Mean	Standard Deviation	Correlation
p	0.2464	0.0174	-0.23
q	0.1731	0.0149	

The structural distribution for p and q in community D is normal:

	Mean	Standard Deviation	Correlation
p	0.1902	0.0154	-0.23
q	0.2357	0.0169	

The composite multinomial model can arise without the specialized structure involving a simpler quantity θ. For $i = 1, \ldots, k$, let $(x_{1i}, \ldots, x_{ri})$ be multinomial with total frequency n_i and quantity $(p_{1i}, \ldots, p_{ri})$. Consider a test of significance for the equality of the multinomial probability vectors:

$$(p_{11}, \ldots, p_{r1}) = \cdots = (p_{1k}, \ldots, p_{rk}).$$

First consider the general model that allows different probability vectors in the different component multinomials. In each component the location quantity can be fitted exactly to the measurement vector; the (ji)th coordinate for the fitted location vector is $\sqrt{x_{ji}}$.

Now consider the restricted model having the same probability vector for each component multinomial. Points with the same probability function should be combined; accordingly the component multinomials should be combined into a single multinomial. Let $m_j = \Sigma_i x_{ji}$; let p_j designate the corresponding probability; and let $n = \Sigma n_i$. The location quantity for the single combined multinomial can be fitted exactly to the measurement vector; the jth coordinate for the fitted location vector is $\sqrt{m_j}$. In order to compare this fitted location vector with the vector for the general model it is necessary to reexpress it in R^{rk}: The (ji)th coordinate for the location vector is then

$$\sqrt{n_i\left(\frac{m_j}{n}\right)} = \left(\frac{\sum_{j'} x_{j'i} \sum_{i'} x_{ji'}}{n}\right)^{1/2}.$$

The squared length of the difference vector is

$$\sum_{ji}\left(\sqrt{x_{ji}} - \sqrt{\frac{n_i m_j}{n}}\right)^2;$$

the corresponding dimension is

$$(r-1)k - (r-1) = (r-1)(k-1).$$

The chi-square value

$$\chi^2 = 4\sum_{ji}\left(\sqrt{x_{ji}} - \sqrt{\frac{n_i m_j}{n}}\right)^2$$

can be compared with the chi-square distribution on $(r-1)(k-1)$ degrees of freedom and the hypothesis of equal probability vectors assessed accordingly.

If the component multinomials are in fact binomials, then the angular transformation mentioned at the end of the preceding section can simplify the analysis.

6 THE HYPERGEOMETRIC MODEL

Consider a single multinomial model with rk outcome events arranged in r rows and k columns. The hypergeometric model arises in a test of the independence of the row and column categories.

Let x_{ji} be the observed frequency for the *cell* in the jth row and ith column. The measurement vector has (ji)th coordinate

$$y_{ji} = \sqrt{x_{ji}}.$$

For the general model allowing unrestricted probabilities for the cells the location quantity can be fitted exactly to the measurement vector. The fitted location vector has (ji)th coordinate

$$y_{ji} = \sqrt{x_{ji}}.$$

For the restricted model with independence the probability for the (ji)th cell is $p_{ji} = p_j^{(1)} p_i^{(2)}$ where $p_j^{(1)}$ is the probability for the event of the jth row and $p_i^{(2)}$ is the probability for the event of the ith column. With independence the row totals provide frequencies for the row probabilities, and the column totals provide frequencies for the column probabilities. The fitted quantity corresponding to rows is

$$\frac{m_j}{n} = \frac{\sum_i x_{ji}}{n} \quad \text{for } p_j^{(1)}.$$

The fitted quantity corresponding to columns is

$$\frac{n_i}{n} = \frac{\sum_j x_{ji}}{n} \quad \text{for } p_i^{(2)}.$$

The difference vector corresponding to the withdrawal from general model to the restricted model has (ji)th coordinate

$$\sqrt{x_{ji}} - \sqrt{\frac{m_j n_i}{n}}.$$

The corresponding chi-square value is

$$\chi^2 = 4\sum_{ji}\left(\sqrt{x_{ji}} - \sqrt{\frac{m_j n_i}{n}}\right)^2$$

This value can be compared with the chi-square distribution on

$$rk - 1 - (r - 1) - (k - 1) = (r - 1)(k - 1)$$

degrees of freedom, and the hypothesis of independence can be assessed accordingly.

NOTES AND REFERENCES

The traditional analysis of frequency data is based on a *chi-square* measure proposed by Karl Pearson (1900):

$$\chi^2 = \sum \frac{(\text{observed frequency} - \text{fitted mean})^2}{\text{fitted mean}}.$$

Some detailed discussion concerning the fitted mean and the appropriate degrees of freedom may be found in Fisher (1958, 1959) and Rao (1965).

The approximate measurement model in this chapter leads to a chi-square measure having the form

$$\chi^2 = 4\sum(\sqrt{\text{observed}} - \sqrt{\text{fitted mean}})^2.$$

These measures can be related to the Poisson distribution with variable x and quantity θ. The Karl Pearson measure derives from the approximate normality of

$$z = \frac{x - \theta}{\sqrt{\theta}}.$$

The measure in this chapter derives from the approximate normality of

$$e^* = 2(\sqrt{x} - \sqrt{\theta}).$$

The variables e^* and z were compared in detail in Chapter Six. The variable e^* presents the frequency in a location relationship to the quantity; this is essential in frequency applications in which an increase in the quantity produces an increase in the frequency. In addition, the variable e^* approaches normality more rapidly over most of its range.

The chi-square measure in this chapter has additional advantages for more complex models: The various chi-square values can be exhibited in an analysis-of-variance table, and the calculations can be based on the methods for the simple regression model. The values of the Karl Pearson measure at various fitted values cannot be compared directly; and the fitted vectors cannot be compared simultaneously as vectors in a Euclidean space.

The location and normal properties of the square root transformation were derived for x and θ large. The typical application involves moderate or large values, and the normal approximation is, in fact, quite accurate. In certain applications, however, there may be x arrays that contain one or more extreme values $x = 0$; for example, the tests of independence in Section 6. The normal approximation can remain reasonably accurate by using $x + \frac{1}{2}$ and $\theta + \frac{1}{2}$ in place of x and θ; for example,

$$\chi^2 = 4\sum_{ji}\left(\sqrt{x_{ji} + \tfrac{1}{2}} - \sqrt{\frac{m_j n_i}{n} + \frac{1}{2}}\right)^2.$$

The data in Section 4 were given by Carver (1927) and analyzed in the traditional chi-square manner by Fisher (1958, 1959). The data in Section 5 were given by Rao (1961), and analyzed in the traditional chi-square manner by Rao (1965).

The examples in this chapter were analyzed by L. M. Steinberg.

Carver, W. A. (1927), A genetic study of certain chlorophyll deficiencies in maize, *Genetics*, **12**, 415–440.

Fisher, R. A. (1958), *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

Fisher, R. A. (1959), *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh.

Pearson, K. (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Phil. Mag*., **1**, 157–175.

Rao, C. R. (1961), A study of large sample test criteria through properties of efficient estimates, *Sankhyā*, **A23**, 25–40.

Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*, Wiley, New York.

PROBLEMS

1. Show that the composite multinomial model is obtained if batches of Poisson variables are conditioned (details in Section 1).

2. Show that the additional conditions applied to the special composite multinomial produce the hypergeometric model (details in Section 1).

3. A die was tossed 1600 times:

Event	1	2	3	4	5	6
Frequency	301	308	340	214	196	241.

Make a test of significance for the hypothesis that the die is true (symmetrical).

4. The progeny of a mating were classified by attribute into three groups:

Event	E_1	E_2	E_3
Frequency	10	53	46.

According to a model the corresponding probabilities should be:

Event	E_1	E_2	E_3
Probability	p^2	$2p(1-p)$	$(1-p)^2$.

where $0 \le p \le 1$. Test whether the data are in accord with the model. If appropriate, determine the structural distribution for p. Start: $p = 0.1$. (Mood and Graybill.)

5. One hundred plants were classified according to two attributes: *large* L or *small* l; *white* W or *colored* w. The frequencies are

	W	w	Totals
L	40	20	60
l	15	25	40
Totals	55	45	100

Analyze the data with the succession of models:

(i) Independence between the attributes (use marginal probability p_1 for L and p_2 for W; start: $p_1 = 0.1$, $p_2 = 0.1$).
(ii) Equal probabilities ($\frac{1}{4}$) for the four cells.

Calculate the analysis-of-variance table; make appropriate tests of significance. (Lindley.)

CHAPTER EIGHT

Inference from Likelihood

Without any structuring relationship between the quantity and the response there remains, in the case of a continuous response, only the likelihood function to identify a response value. This chapter considers multiple observations from such a continuous response. Subject to some regularity conditions, it is shown that, as the number of observations approaches infinity, the likelihood function approaches a limiting form that has a single variable in location relationship to the quantity.

*1 LIKELIHOOD FUNCTION: FAR FROM THE QUANTITY

Consider a continuous response variable x and a quantity θ. Suppose there is no structuring relationship between the quantity and the response, only a probability density function $f(x:\theta)$ for the response variable for each value of the quantity: *the classical model of statistics*. The probability density function is with respect to a differential for x; this is made explicit as needed.

The likelihood function from an observed response value was defined in Section 1, Chapter Four. The alternative form as a log-likelihood function is convenient for analysis here:

$$l(x_0:\theta) = R(x_0) + \ln f(x_0:\theta).$$

This chapter is concerned with the shape of the likelihood function, and it investigates the shape by examining differences for different θ values:

$$l(x_0:\theta'') - l(x_0:\theta') = \ln f(x_0:\theta'') - \ln f(x_0:\theta').$$

(See Figure 1.) For this it is convenient, *in this chapter*, to let $l(x:\theta)$ designate *the logarithm of the density function*,

$$l(x:\theta) = \ln f(x:\theta),$$

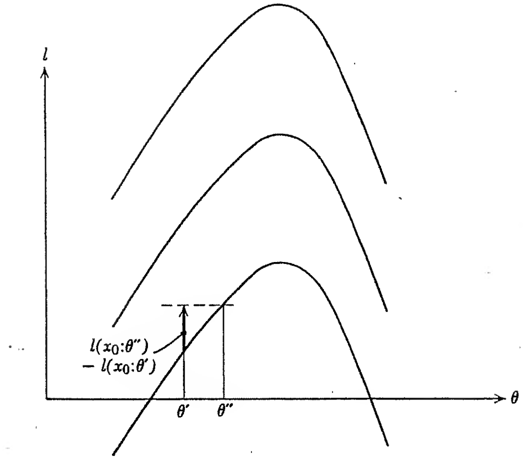


Figure 1 The log-likelihood difference from θ' to θ''.

and to be aware that only *differences*,

$$l(x:\theta'') - l(x:\theta'),$$

represent characteristics of the proper log-likelihood function for a response value x.

In this chapter the likelihood function is examined as a *variable*, as a function of the response *variable* x. For this it is convenient notationally to let θ^0 designate the actual value of the quantity, the value that determines the distribution of the variable x, and to let θ be a free variable designating possible values for the quantity. The likelihood function is examined as a *variable* by analyzing differences, such as

$$\begin{aligned} d(x:\theta) &= l(x:\theta) - l(x:\theta^0) \\ &= \ln f(x:\theta) - \ln f(x:\theta^0), \end{aligned}$$

as *variables*, based on the response *variable* x.

Consider multiple observations on the response: $x_1, \ldots, x_n$. The likelihood difference for the vector response is

$$\begin{aligned} d(\mathbf{x}:\theta) &= l(\mathbf{x}:\theta) - l(\mathbf{x}:\theta^0) \\ &= \ln f(\mathbf{x}:\theta) - \ln f(\mathbf{x}:\theta^0 \\ &= \sum_1^n (l(x_i:\theta) - l(x_i:\theta^0)) \\ &= \sum_1^n d(x_i:\theta). \end{aligned}$$

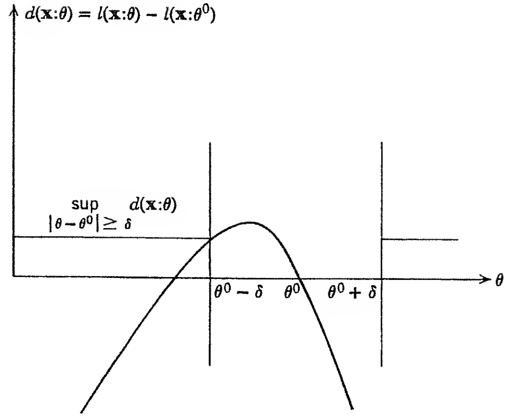


Figure 2 The log-likelihood difference from θ^0 to $\theta : d(\mathbf{x}:\theta)$. The supremum of the log-likelihood difference for θ outside the δ neighborhood of θ^0.

In this section this difference is examined for θ values *outside* a small neighborhood of the actual θ^0. It is shown under mild conditions that the *maximum value outside a given small neighborhood goes to* $-\infty$ *with probability* 1 *as* $n \to \infty$ (see Figure 2). This main result is presented as a theorem in this section; its proof is based on a succession of lemmas.

Lemma 1. If the distribution $f(x:\theta)$ is different from the distribution $f(x:\theta^0)$ and if the mean

$$E\{l(x:\theta^0):\theta^0\}$$

is finite, then

$$E\{l(x:\theta):\theta^0\} < E\{l(x:\theta^0):\theta^0\}$$

or

$$E\{d(x:\theta):\theta^0\} < 0.$$

Proof. A real-valued function $c(t)$ of a real variable t is *strictly convex* if

$$c(at' + (1-a)t'') < ac(t') + (1-a)c(t'')$$

for all t', t'' and $0 < a < 1$. A strictly convex function $c(t)$ has a *line of support* $l(t)$ at any point t':

$$\begin{aligned} l(t) &< c(t), \qquad t \neq t', \\ l(t') &= c(t'), \end{aligned}$$

where $l(t)$ is linear in t. (See Figure 3.) If t is a variable having a distribution with mean $E\{t\} = \nu$, then

$$c(E\{t\}) < E\{c(t)\}$$

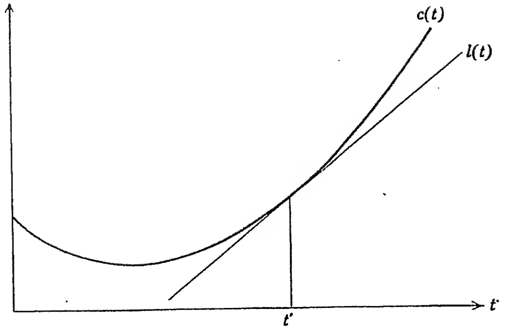


Figure 3 A strictly convex function $c(t)$. A line of support at t'.

unless t has all probability at ν (that is, unless t is a constant). This follows by letting $l(t)$ be the line of support at ν:

$$c(E\{t\}) = l(E\{t\}) = E\{l(t)\} < E\{c(t)\}.$$

Now consider the mean value of the likelihood difference $d(x:\theta) = l(x:\theta) - l(x:\theta^0)$:

$$\begin{aligned} E\{d(x:\theta):\theta^0\} &= E\{l(x:\theta):\theta^0\} - E\{l(x:\theta^0):\theta^0\} \\ &= E\{\ln f(x:\theta):\theta^0\} - E\{\ln f(x:\theta^0):\theta^0\} \\ &= E\left\{\ln \frac{f(x:\theta)}{f(x:\theta^0)}:\theta^0\right\} < \ln E\left\{\frac{f(x:\theta)}{f(x:\theta^0)}:\theta^0\right\} \\ &\leq \ln 1 = 0. \end{aligned}$$

The succession of steps uses the strict convexity of $-\ln t$ and the fact that the integral of $f(x:\theta)$ over points having $f(x:\theta^0) > 0$ is less than or equal to 1. This establishes the lemma.

Lemma 2. If the distribution $f(x:\theta)$ is different from the distribution $f(x:\theta^0)$, and if the mean

$$E\{l(x:\theta^0):\theta^0\}$$

is finite, then for a sequence of response observations, $x_1, x_2, \ldots$,

$$\Pr\left\{\lim_{n\to\infty} \sum_1^n d(x_i:\theta) = -\infty:\theta^0\right\} = 1.$$

Proof. The lemma is concerned with the likelihood difference

$$l(\mathbf{x}:\theta) - l(\mathbf{x}:\theta^0) = \sum_1^n \left(l(x_i:\theta) - l(x_i:\theta^0)\right) = \sum_1^n d(x_i:\theta);$$

the lemma asserts that with probability 1 the limit is $-\infty$ as $n \to \infty$.

By Lemma 1 the mean value of $d(x:\theta)$ satisfies

$$E\{d(x:\theta):\theta^0\} = \epsilon < 0.$$

Then, by the strong law of large numbers,

$$\Pr\left(\lim_{n\to\infty} \frac{\sum_1^n d(x_i:\theta)}{n} = \epsilon:\theta^0\right) = 1.$$

Hence

$$\Pr\left\{\lim_{n\to\infty} \sum_1^n d(x_i:\theta) = -\infty:\theta^0\right\} = 1,$$

and the lemma is established.

Corollary. If $E\{d(x):\theta^0\} < 0$ for some $d(x)$, then

$$\Pr\left\{\lim_{n\to\infty} \sum_1^n d(x_i) = -\infty:\theta^0\right\} = 1.$$

Now consider several θ values, $\theta^1, \ldots, \theta^h$ and suppose that the distribution of x for each of these values is different from the distribution $f(x:\theta^0)$. The following lemma asserts that the maximum likelihood difference to these values goes to $-\infty$ with probability 1 as $n \to \infty$.

Lemma 3. If the distributions $f(x:\theta^1), \ldots, f(x:\theta^h)$ are different from the distribution $f(x:\theta^0)$, and if the mean

$$E\{l(x:\theta^0):\theta^0\}$$

is finite, then for a sequence of response observations $x_1, x_2, \ldots$

$$\Pr\left\{\lim_{n\to\infty} \max_{\alpha=1}^{h} \sum_1^n d(x_i:\theta^\alpha) = -\infty:\theta^0\right\} = 1.$$

Proof. Consider a sequence $s_n(\alpha)$ of real numbers for each $\alpha = 1, \ldots, h$. If

$$\lim_{n\to\infty} s_n(\alpha) = -\infty$$

for each α, then

$$\lim_{n\to\infty} \max_{\alpha=1}^{h} s_n(\alpha) = -\infty;$$

and conversely.

Consider events $A_1, \ldots, A_h$ and suppose $\Pr\{A_\alpha\} = 1$ for each α. Then the probability that all the events A_α occur is equal to 1:

$$\Pr\left\{\bigcap_\alpha A_\alpha\right\} = 1 - \Pr\left\{\bigcup_\alpha \bar{A}_\alpha\right\}$$
$$\geq 1 - \sum_{\alpha=1}^{h} \Pr\{\bar{A}_\alpha\} = 1 - \sum_{\alpha=1}^{h} 0$$
$$= 1,$$

where $\bar{A}_\alpha$ designates the nonoccurrence of A_α.

Let A_α be the *event*:

$$\lim_{n\to\infty} \sum_1^n d(x_i:\theta^\alpha) = -\infty.$$

The results in the preceding paragraphs then establish the lemma.

Corollary. If $E\{d^\alpha(x):\theta^0\} < 0$ for $d^1(x), \ldots, d^h(x)$, then

$$\Pr\left\{\lim_{n\to\infty} \max_{\alpha=1}^{h} \sum_1^n d^\alpha(x_i) = -\infty : \theta^0\right\} = 1.$$

The main theorem is concerned with the maximum likelihood difference to θ values outside a neighborhood of θ^0. This needs continuity so that what happens at a θ^α value controls what happens for θ values near θ^α. It also needs some sort of uniformity so that a finite number of θ^α values controls what happens for all the values outside a neighborhood of θ^0.

Now suppose that the range for the quantity θ is Euclidean space R^k. The assumptions needed for the main theorem can be expressed more easily if the point-at-infinity ∞ is added to R^k. Let $d(\theta', \theta'')$ designate Euclidean distance in R^k, let

$$B_\rho(\theta') = \{\theta: \quad d(\theta, \theta') < \rho\}$$

designate the ball of radius ρ about θ', and let

$$B_\rho(\infty) = \{\theta: \quad d(\theta, 0) > 1/\rho\}$$

designate a ball about ∞ (the region outside the sphere of radius $1/\rho$). The balls $B_\rho(\theta)$ are neighborhoods for the points in $R^k \cup \{\infty\}$.

Assumption 1. $f(x:\theta)$ is a continuous function† of θ in $R^k \cup \{\infty\}$. For each $\theta \neq \theta^0$, the distribution $f(x:\theta)$ is different from the distribution

† The function $f(x:\infty)$ is taken to be the limiting function $\lim_{\theta\to\infty} f(x:\theta)$ as implied by continuity.

$f(x:\theta^0)$. For† each θ' in $R^k \cup \{\infty\}$ there is a neighborhood $B_\rho(\theta')$ such that

$$\sup_{\theta \text{ in } B_\rho(\theta')} l(x:\theta) \leq M(x:\theta');$$

the mean values

$$E\{M(x:\theta'):\theta^0\}, \qquad E\{l(x:\theta^0):\theta^0\}:$$

are finite.

Lemma 4. If $f(x:\theta)$ satisfies Assumption 1, then

$$\lim_{\rho\to 0} E\left\{\sup_{\theta \text{ in } B_\rho(\theta')} l(x:\theta):\theta^0\right\} = E\{l(x:\theta'):\theta^0\}.$$

Proof. By the continuity part of Assumption 1, it follows that

$$\lim_{\rho\to 0} \sup_{\theta \text{ in } B_\rho(\theta')} l(x:\theta) = l(x:\theta').$$

The function

$$\sup_{\theta \text{ in } B_\rho(\theta')} l(x:\theta)$$

is monotone decreasing as $\rho \to 0$. For ρ small enough it is bounded above by the function $M(x:\theta')$, which has finite mean value. It follows, by the monotone convergence theorem for integrals, that the limit operation with respect to ρ can be carried outside the integral sign:

$$E\left\{\lim_{\rho\to 0} \sup_{\theta \text{ in } B_\rho(\theta')} l(x:\theta):\theta^0\right\} = \lim_{\rho\to 0} E\left\{\sup_{\theta \text{ in } B_\rho(\theta')} l(x:\theta):\theta^0\right\}.$$

This establishes the lemma.

Theorem 5. If the classical model $f(x:\theta)$ satisfies Assumption 1, then

$$\Pr\left\{\lim_{n\to\infty} \sup_{d(\theta,\theta^0)\geq\delta} \sum_1^n d(x_i:\theta) = -\infty:\theta^0\right\} = 1.$$

Proof. For each value θ' different from θ^0 there is, by Lemmas 4 and 1, a neighborhood $B_\rho(\theta')$ such that

$$E\left\{\sup_{\theta \text{ in } B_\rho(\theta')} l(x:\theta):\theta^0\right\} < E\{l(x:\theta^0):\theta^0\}.$$

† This part of Assumption 1 can be replaced by the simpler but stronger condition that $|\partial l(x:\theta)/\partial\theta| \leq M(x)$ (all θ) with $E\{M(x):\theta^0\}$ finite.

Consider the region of θ values in $R^k \cup \{\infty\}$ having $d(\theta, \theta^0) \geq \delta$: each θ' value in this region has a neighborhood $B_\rho(\theta')$ for which the preceding inequality holds. By the Heine-Borel theorem a finite number of these neighborhoods can be found that cover the region. Let the corresponding θ values be $\theta^1, \ldots, \theta^h$ (one of these is $\theta = \infty$). Then by the corollary of Lemma 3, with

$$d_\alpha(x) = \sup_{\theta \text{ in } B_\rho(\theta^\alpha)} l(x:\theta) - l(x:\theta^0),$$

it follows that

$$\Pr\left\{\lim_{n\to\infty} \max_{\alpha=1}^{h} \sum_{i=1}^{n} \left(\sup_{\theta \text{ in } B_\rho(\theta^\alpha)} l(x_i:\theta) - l(x_i:\theta^0)\right) = -\infty : \theta^0\right\} = 1.$$

The theorem is then established by noting that

$$\max_\alpha \sum_{i=1}^{n} \left(\sup_{\theta \text{ in } B_\rho(\theta^\alpha)} l(x_i:\theta) - l(x_i:\theta^0)\right) \geq \sup_{d(\theta,\theta^0)\geq\delta} \sum_{i=1}^{n} (l(x_i:\theta) - l(x_i:\theta^0)).$$

Under the mild conditions of Assumption 1, the theorem asserts that for a sequence of response observations $x_1, x_2, \ldots$ the maximum likelihood (as a difference relative to θ^0) outside a neighborhood of the actual θ^0 goes to $-\infty$ with probability 1.

2 LIKELIHOOD FUNCTION: NEAR THE QUANTITY

Now consider the form of the likelihood function *near the quantity*, in a small neighborhood of the actual θ^0. With multiple observations $x_1, \ldots, x_n$ on the response, the likelihood difference from θ^0 to a general value θ is

$$\begin{aligned} d(\mathbf{x}:\theta) &= l(\mathbf{x}:\theta) - l(\mathbf{x}:\theta^0) \\ &= \ln f(\mathbf{x}:\theta) - \ln f(\mathbf{x}:\theta^0) \\ &= \sum_1^n (l(x_i:\theta) - l(x_i:\theta^0)) = \sum_1^n d(x_i:\theta). \end{aligned}$$

This likelihood difference is examined in this section for θ in a small neighborhood of θ^0; and it is examined as a *variable*, based on the multiple response *variable* $\mathbf{x}$. It is shown that as $n \to \infty$ the limiting form is quadratic, with a single variable in location relationship to the quantity. The analysis is given in Section 2.1 for the notationally easy case of a real quantity θ; the analogous results for a vector quantity are summarized at the end of Section 2.2.

2.1 A Real-Valued Quantity θ.

Consider first the slope of the likelihood function at the actual value θ^0. Let

$$l^{(1)}(x:\theta) = \frac{\partial}{\partial\theta} l(x:\theta) = \frac{\partial}{\partial\theta} \ln f(x:\theta),$$

$$l^{(11)}(x:\theta) = \frac{\partial^2}{\partial\theta^2} l(x:\theta) = \frac{\partial^2}{\partial\theta^2} \ln f(x:\theta),$$

$$f^{(1)}(x:\theta) = \frac{\partial}{\partial\theta} f(x:\theta),$$

$$f^{(11)}(x:\theta) = \frac{\partial^2}{\partial\theta^2} f(x:\theta).$$

The following assumptions are convenient.

Assumption 2. In a neighborhood of θ^0,

$$|f^{(11)}(x:\theta)| \leq M_1(x)$$

and $M_1(x)$ and $f^{(1)}(x:\theta^0)$ are integrable.

Assumption 3. The likelihood derivatives $l^{(1)}(x:\theta)$, $l^{(11)}(x:\theta)$ exist in a neighborhood of θ^0 and

$$E\{l^{(11)}(x:\theta^0):\theta^0\}$$

is finite-valued.

Lemma 6 establishes mean value properties of the likelihood derivatives:

Lemma 6. If the classical model $f(x:\theta)$ satisfies Assumptions 2 and 3, then

$$E\{l^{(1)}(x:\theta^0):\theta^0\} = 0,$$

$$E\{-l^{(11)}(x:\theta^0):\theta^0\} = j(\theta^0) = \operatorname{var}\{l^{(1)}(x:\theta^0):\theta^0\}.$$

Proof. The density function $f(x:\theta)$ can be expanded in a Taylor series with respect to θ at θ^0:

$$f(x:\theta) = f(x:\theta^0) + (\theta - \theta^0) f^{(1)}(x:\theta^0) + \frac{(\theta - \theta^0)^2}{2!} f^{(11)}(x:\theta^*),$$

where $|\theta^* - \theta^0| \leq |\theta - \theta^0|$. By Assumption 2 it follows that $f(x:\theta)$ and $f^{(1)}(x:\theta)$ are bounded by integrable functions for θ in a neighborhood of θ^0. This permits differentiation with respect to θ to be carried through the

integral sign:

$$\int f(x:\theta)\,dx = 1,$$

$$\int f^{(1)}(x:\theta)\,dx = 0,$$

$$\int f^{(11)}(x:\theta)\,dx = 0$$

(the differential dx can be a Euclidean differential, or it can be a general differential). By Assumption 3 the integrands can be rearranged:

$$\int f^{(1)}(x:\theta)\,dx = \int l^{(1)}(x:\theta)f(x:\theta)\,dx = 0,$$

$$\int f^{(11)}(x:\theta)\,dx = \int \big(l^{(11)}(x:\theta)f(x:\theta) + l^{(1)}(x:\theta)f^{(1)}(x:\theta)\big)\,dx = 0.$$

At θ^0 this gives

$$E\{l^{(1)}(x:\theta^0):\theta^0\} = 0,$$

$$E\{-l^{(11)}(x:\theta^0):\theta^0\} = E\{(l^{(1)}(x:\theta^0))^2:\theta^0\},$$

which establishes the lemma. The mean-value characteristic of the likelihood slope at θ^0,

$$E\{(l^{(1)}(x:\theta^0))^2:\theta^0\} = \operatorname{var}\{l^{(1)}(x:\theta^0):\theta^0\} = j(\theta^0),$$

is called the *Fisher information* at θ^0.

For a sequence of response observations $x_1, x_2, \ldots$, the second lemma in this section (Lemma 7) establishes some distribution properties of the likelihood derivative at θ^0. For the vector $\mathbf{x} = (x_1, \ldots, x_n)'$ the logarithm of the density function is

$$l(\mathbf{x}:\theta) = \sum_1^n l(x_i:\theta),$$

and derivatives are

$$l^{(1)}(\mathbf{x}:\theta) = \sum_1^n l^{(1)}(x_i:\theta),$$

$$l^{(11)}(\mathbf{x}:\theta) = \sum_1^n l^{(11)}(x_i:\theta).$$

At θ^0 the properties of mean and variance of independent variables give

$$E\{l^{(1)}(\mathbf{x}:\theta^0):\theta^0\} = nE\{l^{(1)}(x:\theta^0):\theta^0\},$$

$$\operatorname{var}\{l^{(1)}(\mathbf{x}:\theta^0):\theta^0\} = n\operatorname{var}\{l^{(1)}(x:\theta^0):\theta^0\}.$$

Lemma 7. If $E\{l^{(1)}(x:\theta^0):\theta^0\}$, $\operatorname{var}\{l^{(1)}(x:\theta^0):\theta^0\}$ exist and if

$$E\{l^{(1)}(x:\theta^0):\theta^0\} = 0,$$

then the distribution of

$$\frac{l^{(1)}(\mathbf{x}:\theta^0)}{\sqrt{n}} = \frac{\sum_1^n l^{(1)}(x_i:\theta^0)}{\sqrt{n}}$$

approaches the normal distribution with mean 0 and variance $j(\theta^0)$ as $n \to \infty$. If $j(\theta^0) > 0$, then the distribution of

$$w = \frac{l^{(1)}(\mathbf{x}:\theta^0)}{\sqrt{n}\,j(\theta^0)} = \frac{\sum_1^n l^{(1)}(x_i:\theta^0)}{\sqrt{n}\,j(\theta^0)}$$

approaches the normal distribution with mean 0 and variance $1/j(\theta^0)$.

Proof. A direct application of the central limit theorem.

Now consider the form of the likelihood function near θ^0 as the number n of response observations becomes large. Suppose the model $f(x:\theta)$ satisfies Assumptions 2 and 4 (Assumption 4 is a needed stronger version of Assumption 3):

Assumption 4. In a neighborhood of θ^0,

$$|l^{(111)}(x:\theta)| < M_2(x),$$

and

$$E\{l(x:\theta^0):\theta^0\}, \qquad E\{l^{(1)}(x:\theta^0):\theta^0\}$$

$$E\{l^{(11)}(x:\theta^0):\theta^0\} \neq 0, \qquad E\{M_2(x):\theta^0\} = -q(\theta^0)$$

are finite-valued.

The likelihood difference near θ^0 can be expanded in a Taylor series (Assumption 4):

$$l(\mathbf{x}:\theta) - l(\mathbf{x}:\theta^0)$$

$$= (\theta-\theta^0)l^{(1)}(\mathbf{x}:\theta^0) + \frac{(\theta-\theta^0)^2}{2!}l^{(11)}(\mathbf{x}:\theta^0) + \frac{(\theta-\theta^0)^3}{3!}RM_2(\mathbf{x}),$$

where

$$M_2(\mathbf{x}) = \sum_1^n M_2(x_i)$$

and $|R| \leq 1$; the expansion replaces a continuum of functions of $\mathbf{x}$ indexed by θ (on the left side) by effectively *three* functions of $\mathbf{x}$ (on the right side).

With Lemma 6 the strong law of large numbers gives

$$\Pr\left\{\lim_{n\to\infty}\frac{l^{(11)}(\mathbf{x}:\theta^0)}{n} = -j(\theta^0):\theta^0\right\} = 1,$$

$$\Pr\left\{\lim_{n\to\infty}\frac{M_2(\mathbf{x})}{n} = -q(\theta^0):\theta^0\right\} = 1.$$

The convergence theorem for orthogonal variables gives

$$\Pr\left\{\lim_{n\to\infty}\frac{l^{(1)}(\mathbf{x}:\theta^0)}{n^{\frac{1}{2}+\epsilon}} = 0:\theta^0\right\} = 1$$

for any $\epsilon > 0$. These probability limits suggest a rearrangement of the expression for the likelihood difference near θ^0. Let

$$V(\mathbf{x}:\theta) = -\left(l^{(11)}(\mathbf{x}:\theta^0) + \frac{\theta-\theta^0}{3}RM_2(\mathbf{x})\right);$$

then

$$\begin{aligned} l(\mathbf{x}:\theta) - l(\mathbf{x}:\theta^0) &= (\theta-\theta^0)l^{(1)}(\mathbf{x}:\theta^0) - (\theta-\theta^0)^2\frac{V(\mathbf{x}:\theta)}{2} \\ &= -\tfrac{1}{2}V(\mathbf{x}:\theta)\left(\theta-\theta^0-\frac{l^{(1)}(\mathbf{x}:\theta^0)}{V(\mathbf{x}:\theta)}\right)^2 + \frac{1}{2}\frac{(l^{(1)}(\mathbf{x}:\theta^0))^2}{V(\mathbf{x}:\theta)} \\ &= -\frac{1}{2}\frac{V(\mathbf{x}:\theta)}{n}\left(\sqrt{n}(\theta-\theta^0) - \frac{l^{(1)}(\mathbf{x}:\theta^0)/\sqrt{n}}{V(\mathbf{x}:\theta)/n}\right)^2 \\ &\quad + \frac{1}{2}\frac{(l^{(1)}(\mathbf{x}:\theta^0)/\sqrt{n})^2}{V(\mathbf{x}:\theta)/n}. \end{aligned}$$

Consider the likelihood difference about θ^0 in units of length $n^{-\frac{1}{2}}$:

$$\theta = \theta^0 + \tau n^{-\frac{1}{2}}.$$

The likelihood difference is

$$\begin{aligned} &l(\mathbf{x}:\theta^0+\tau n^{-\frac{1}{2}}) - l(\mathbf{x}:\theta^0) \\ &\quad = -\frac{1}{2}\frac{V(\mathbf{x}:\theta)}{n}\left(\tau - w\left(\frac{j(\theta^0)}{V(\mathbf{x}:\theta)/n}\right)\right)^2 + \frac{j(\theta^0)}{2}w^2\cdot\left(\frac{j(\theta^0)}{V(\mathbf{x}:\theta)/n}\right), \end{aligned}$$

where

$$w = \frac{l^{(1)}(\mathbf{x}:\theta^0)}{\sqrt{n}\,j(\theta^0)}.$$

Preceding results derived from the strong law of large numbers give

$$\Pr\left\{\lim_{n\to\infty}\frac{V(\mathbf{x}:\theta)}{n} = j(\theta^0):\theta^0\right\} = 1,$$

and hence

$$\Pr\left\{\lim_{n\to\infty}\frac{V(\mathbf{x}:\theta)/n}{j(\theta^0)} = 1:\theta^0\right\} = 1.$$

By Lemma 7, the variable w has a limiting normal distribution with mean 0 and variance $1/j(\theta^0)$. It follows then that *with probability* 1 *the limiting form of the likelihood difference for bounded* τ *is*

$$\lim_{n\to\infty}(l(\mathbf{x}:\theta^0+\tau n^{-\frac{1}{2}}) - l(\mathbf{x}:\theta^0)) = -\frac{j(\theta^0)}{2}(w-\tau)^2 + \frac{j(\theta^0)}{2}w^2.$$

The limiting form involves a single variable w *having a limiting normal distribution with mean* 0 *and variance* $1/j(\theta^0)$.

Now consider the likelihood difference *elsewhere* in the neighborhood $(\theta^0-\delta, \theta^0+\delta)$ of θ^0. For this choose δ small enough that Assumption 4 will hold in the interval $(\theta^0-\delta, \theta^0+\delta)$ and small enough that

$$j(\theta^0) - \frac{\delta}{3}|q(\theta^0)| > V,$$

where V is a positive number. Let

$$\theta = \theta^0 + \tau n^{-\frac{1}{2}+\epsilon},$$

where $|\tau| < \delta$ and $0 < \epsilon < \frac{1}{2}$. The likelihood difference is

$$\begin{aligned} l(\mathbf{x}:\theta^0+\tau n^{-\frac{1}{2}+\epsilon}) - l(\mathbf{x}:\theta^0) &= \tau n^{-\frac{1}{2}+\epsilon}l^{(1)}(\mathbf{x}:\theta^0) - \tau^2 n^{-1+2\epsilon}\frac{V(\mathbf{x}:\theta)}{2} \\ &= -n^{2\epsilon}\left(\frac{\tau^2}{2}\frac{V(\mathbf{x}:\theta)}{n} - \frac{\tau}{n^{\epsilon-\epsilon_0}}\frac{l^{(1)}(\mathbf{x}:\theta^0)}{n^{\frac{1}{2}+\epsilon_0}}\right). \end{aligned}$$

With probability 1 the expression in parentheses is greater than $\tau^2V/2$ for all ϵ in the range $0 < \epsilon_0 \le \epsilon \le \frac{1}{2}$. Then *with probability* 1 *the maximum likelihood difference for* ϵ *in the range* $0 < \epsilon_0 \le \epsilon \le \frac{1}{2}$ *has limit* $-\infty$. This holds for all $\epsilon_0 (0 < \epsilon_0 \le \frac{1}{2})$.

The maximum likelihood difference *outside* a neighborhood $(\theta^0-\delta, \theta^0+\delta)$, by Assumption 1 and Theorem 5 (preceding section), goes to $-\infty$ with probability 1.

Hence; *the limiting form of the likelihood difference relative to the actual* θ^0 *is*

$$-\frac{j(\theta^0)}{2}(w-\tau)^2 + \frac{j(\theta^0)}{2}w^2$$

in terms of τ *given by* $\theta = \theta^0 + \tau n^{-\frac{1}{2}}$, *and is* $-\infty$ *otherwise; the variable* w *has a limiting normal distribution with mean* 0 *and variance* $1/j(\theta^0)$ (see Figure 4).

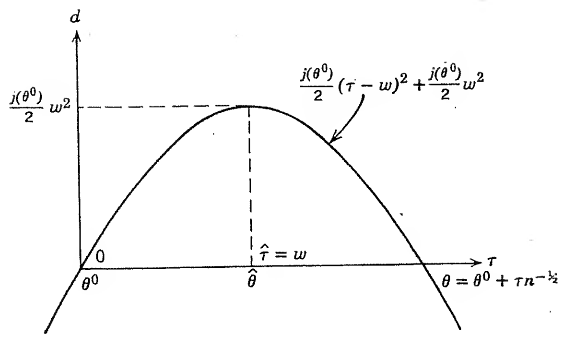


Figure 4. The limiting form of the likelihood difference relative to the actual θ^0.

The likelihood function has a maximum at $\tau = w$. If the likelihood difference is taken with respect to the maximum likelihood, the limiting form of the likelihood function becomes

$$-\frac{j(\theta^0)}{2}(w-\tau)^2 + \frac{j(\theta^0)}{2}w^2 - \left\{-\frac{j(\theta^0)}{2}0^2 + \frac{j(\theta^0)}{2}w^2\right\} = -\frac{j(\theta^0)}{2}(w-\tau)^2$$

(see Figure 5).

The limiting form of the likelihood function involves a single variable w; the variable w has a limiting normal distribution with variance $1/j(\theta^0)$; the

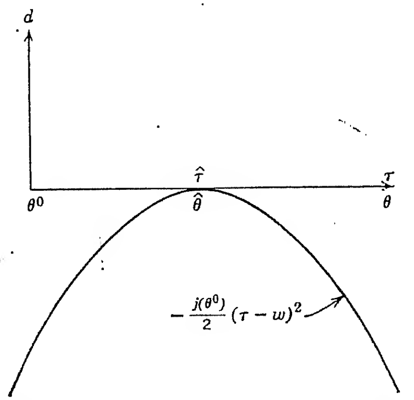


Figure 5. The limiting form of the likelihood difference taken with respect to the maximum likelihood (at $\tau = w$).

limiting form of the likelihood function is the likelihood function of w as a normal variable with location quantity τ; the *actual value* for the quantity τ is $\tau = 0$ (note that the actual value for the quantity was chosen as the origin for the present analysis).

***2.2 A Vector-Valued Quantity $\boldsymbol{\theta}$.** Consider now a vector-valued quantity $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$; and let $\boldsymbol{\theta}^0 = (\theta_1^0, \ldots, \theta_k^0)$ designate the actual value of the quantity, the value that determines the distribution of the response variable x.

Derivatives of the likelihood function can be taken with respect to the different coordinates of $\boldsymbol{\theta}$:

$$\frac{\partial}{\partial\theta_j} l(x:\boldsymbol{\theta}) = l^{(j)}(x:\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j}\ln f(x:\boldsymbol{\theta}),$$
$$\mathbf{l}^D(x:\boldsymbol{\theta}) = (l^{(1)}(x:\boldsymbol{\theta}), \ldots, l^{(k)}(x:\boldsymbol{\theta}))',$$
$$\frac{\partial^2}{\partial\theta_j\,\partial\theta_{j'}} l(x:\boldsymbol{\theta}) = l^{(jj')}(x:\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_j\,\partial\theta_{j'}}\ln f(x:\boldsymbol{\theta}),$$
$$\frac{\partial}{\partial\theta_j} f(x:\boldsymbol{\theta}) = f^{(j)}(x:\boldsymbol{\theta}),$$
$$\frac{\partial^2}{\partial\theta_j\,\partial\theta_{j'}} f(x:\boldsymbol{\theta}) = f^{(jj')}(x:\boldsymbol{\theta}).$$

Assumptions 2, 3, and 4 can be generalized by replacing a first derivative by each of the first-order derivatives in turn; by replacing a second derivative by each of the second-order derivatives in turn; by replacing a third derivative by each of the third-order derivatives in turn; and by replacing

$$E\{l^{(11)}(x:\theta^0):\theta^0\} \neq 0$$

by the nonsingularity of the matrix

$$E\left\{\begin{bmatrix} l^{(11)}(x:\boldsymbol{\theta}^0) & \cdots & l^{(1k)}(x:\boldsymbol{\theta}^0) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ l^{(k1)}(x:\boldsymbol{\theta}^0) & \cdots & l^{(kk)}(x:\boldsymbol{\theta}^0) \end{bmatrix} : \boldsymbol{\theta}^0\right\}.$$

For a vector-valued quantity $\boldsymbol{\theta}$, Lemma 6 becomes the following:

Lemma 8. If the classical model $f(x:\boldsymbol{\theta})$ satisfies the generalized Assumptions 2, and 3, then

$$E\{\mathbf{l}^D(x:\boldsymbol{\theta}^0):\boldsymbol{\theta}^0\} = \mathbf{0},$$
$$E\{\mathbf{l}^D(x:\boldsymbol{\theta}^0)\mathbf{l}'^D(x:\boldsymbol{\theta}^0):\boldsymbol{\theta}^0\} = J(\boldsymbol{\theta}^0),$$

where

$$J(\theta^0) = \begin{bmatrix} j_{11}(\theta^0) & \cdots & j_{1k}(\theta^0) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ j_{k1}(\theta^0) & \cdots & j_{kk}(\theta^0) \end{bmatrix},$$

$$j_{jj'}(\theta^0) = E\{-l^{(jj')}(x:\theta^0):\theta^0\} = \text{cov}\,\{l^{(j)}(x:\theta^0), l^{(j')}(x:\theta^0):\theta^0\}.$$

Proof. A direct extension of the proof of Lemma 6. The mean-value characteristic of the likelihood gradient $(\mathbf{l}^D(x:\theta))$ at θ^0,

$$J(\theta^0) = \text{cov}\,\{\mathbf{l}^D(x:\theta^0), \mathbf{l}^D(x:\theta^0):\theta^0\},$$

is called the *Fisher information matrix* at θ^0.

For a sequence of response observations $x_1, x_2, \ldots$ the likelihood derivatives are additive:

$$l(\mathbf{x}:\theta) = \sum_1^n l(x_i:\theta),$$

$$l^{(j)}(\mathbf{x}:\theta) = \sum_1^n l^{(j)}(x_i:\theta),$$

$$l^{(jj')}(\mathbf{x}:\theta) = \sum_1^n l^{(jj')}(x_i:\theta).$$

At θ^0 the properties of mean and variance for independent variables give

$$E\{\mathbf{l}^D(\mathbf{x}:\theta^0):\theta^0\} = nE\{\mathbf{l}^D(x:\theta^0):\theta^0\},$$

$$\text{cov}\,\{\mathbf{l}^D(\mathbf{x}:\theta^0), \mathbf{l}^D(\mathbf{x}:\theta^0):\theta^0\} = n\,\text{cov}\,\{\mathbf{l}^D(x:\theta^0), \mathbf{l}^D(x:\theta^0):\theta^0\}.$$

For the vector-valued quantity θ, Lemma 7 becomes

Lemma 9. If the mean and variance of $\mathbf{l}^D(x:\theta^0)$ exist at θ^0, and if $E\{\mathbf{l}^D(x:\theta^0):\theta^0\} = \mathbf{0}$, then the distribution of

$$\frac{\mathbf{l}^D(\mathbf{x}:\theta^0)}{\sqrt{n}} = \frac{\sum_1^n \mathbf{l}^D(x_i:\theta^0)}{\sqrt{n}}$$

approaches the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $J(\theta^0)$. If $J(\theta^0)$ is nonsingular, then the distribution of

$$\mathbf{w} = \frac{1}{\sqrt{n}} J^{-1}(\theta^0)\mathbf{l}^D(\mathbf{x}:\theta^0) = \frac{1}{\sqrt{n}} J^{-1}(\theta^0) \sum_1^n \mathbf{l}^D(x_i:\theta^0)$$

approaches the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $J^{-1}(\theta^0)$.

Proof. A direct application of the central limit theorem for vector variables.

Now consider the form of the likelihood function near θ^0 as the number n of response observations becomes large. Suppose the model $f(x:\theta)$ satisfies the generalized Assumptions 2 and 4. The likelihood difference near θ^0 can be expanded in a Taylor series and rearranged following the pattern for a real quantity.

The limiting form of the likelihood difference relative to the actual θ^0 *is*

$$-\tfrac{1}{2}(\mathbf{w} - \boldsymbol{\tau})'J(\theta^0)(\mathbf{w} - \boldsymbol{\tau}) + \tfrac{1}{2}\mathbf{w}'J(\theta^0)\mathbf{w}$$

in terms of $\theta = \theta^0 + \boldsymbol{\tau}n^{-1/2}$, *and is* $-\infty$ *otherwise. The variable* $\mathbf{w}$,

$$\mathbf{w} = \frac{1}{\sqrt{n}} J^{-1}(\theta^0)\mathbf{l}^D(\mathbf{x}:\theta^0),$$

has a limiting multivariate normal distribution with means equal to zero and covariance matrix $J^{-1}(\theta^0)$.

The likelihood function has a maximum at $\boldsymbol{\tau} = \mathbf{w}$. If the likelihood difference is taken with respect to the maximum likelihood, then the limiting form of the likelihood function becomes

$$-\tfrac{1}{2}(\mathbf{w} - \boldsymbol{\tau})'J(\theta^0)(\mathbf{w} - \boldsymbol{\tau}) + \tfrac{1}{2}\mathbf{w}'J(\theta^0)\mathbf{w} - [-\tfrac{1}{2}0 + \tfrac{1}{2}\mathbf{w}'J(\theta^0)\mathbf{w}]$$
$$= -\tfrac{1}{2}(\mathbf{w} - \boldsymbol{\tau})'J(\theta^0)(\mathbf{w} - \boldsymbol{\tau}).$$

See Figure 6.

The limiting form of the likelihood function involves a single variable $\mathbf{w}$; the variable $\mathbf{w}$ has a limiting normal distribution with covariance matrix $J^{-1}(\theta^0)$; the limiting form of the likelihood function is that of $\mathbf{w}$ as a normal variable with location quantity $\boldsymbol{\tau}$; the *actual value* for the quantity $\boldsymbol{\tau}$ is $\boldsymbol{\tau} = \mathbf{0}$ (note that the actual value for the quantity was chosen as the origin for the present description).

3 LIKELIHOOD INFERENCE: LARGE SAMPLE

Consider a continuous response variable x and a quantity θ. Suppose there is no structuring relationship between the quantity θ and the response x—just a classical model $f(x:\theta)$ satisfying Assumptions 1, 2, and 4. Consider a large number n of response observations $x_1, \ldots, x_n$.

Within the classical model a multiple response vector $(x_1, \ldots, x_n)$ has only its likelihood function to identify it. By the results in the preceding section, for sufficiently large n the likelihood function has normal quadratic

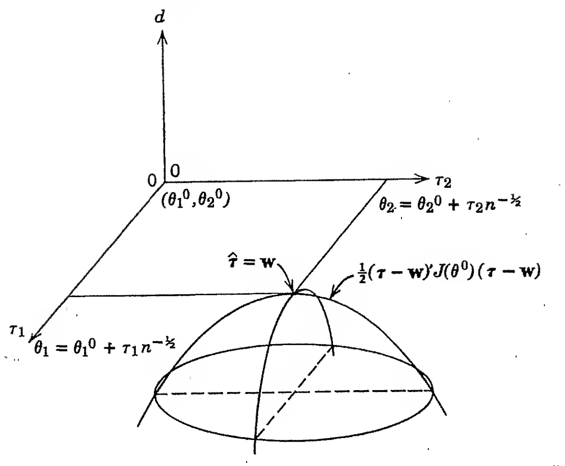


Figure 6. The limiting form of the likelihood difference taken relative to the maximum likelihood; **w** is a variable with limiting multivariate normal distribution with covariance matrix $J^{-1}(\theta^0)$; the location quantity for **w** is $\boldsymbol{\tau}$ and $\boldsymbol{\tau}$ has actual value **0** (but only because the actual value was chosen as the origin in the analysis).

form in the neighborhood of the actual value and approximates $-\infty$ elsewhere. In the neighborhood of a value θ^* near the actual value the likelihood depends on the single variable w:

$$w = \sqrt{n}\,\frac{l^{(1)}(\mathbf{x}:\theta^*)}{nj(\theta^*)}.$$

The variable w has a limiting normal distribution with location quantity† τ,

$$\tau = \sqrt{n}(\theta - \theta^*),$$

and variance $1/j(\theta^*)$; the limiting likelihood function is the likelihood function for the normal variable w.

A change in the quantity θ produces a change in the likelihood function $l(x:\theta)$ at various response values x. This produces a decrease in probability for some values and an increase for other values. For a large sample this produces a loss of certain response observations and a gain of other response observations, and thereby produces a change in w. The large sample model

† If $\theta = \theta^0$ designates the actual value for the *quantity* θ, then $\tau = \tau^0 = \sqrt{n}(\theta^0 - \theta^*)$ designates the actual value of the *location quantity* τ. The details of the change of reference point from θ^0 to an adjacent θ^* are examined in Problem 3.

can then be approximated by the *simple measurement model:*

$$e,$$
$$w = \tau + e.$$

The model has an error variable e with a normal distribution with mean 0 *and variance* $1/j(\theta^*)$; *and it has a structural equation in which a realized value from the error distribution provides the link between the observed value of w,*

$$w = \sqrt{n}\,\frac{l^{(1)}(\mathbf{x}:\theta^*)}{nj(\theta^*)},$$

and the unknown value of τ,

$$\tau = \sqrt{n}(\theta - \theta^*).$$

This measurement model with reference value θ^* is applicable for θ^* close to the true θ^0 (within the range for the approximations in Section 2.1).

For convenience, the measurement model can be expressed directly in terms of the quantity θ. A change of scale by the factor $n^{-1/2}$ gives

$$\frac{e}{\sqrt{n}},$$
$$\frac{l^{(1)}(\mathbf{x}:\theta^*)}{nj(\theta^*)} = (\theta - \theta^*) + \frac{e}{\sqrt{n}}.$$

(Note that the denominator $nj(\theta^*)$ is the variance of the likelihood derivative for the multiple response calculated at θ^*:

$$nj(\theta^*) = \operatorname{var}\{l^{(1)}(\mathbf{x}:\theta^*):\theta^*\} = \operatorname{var}\{\textstyle\sum l^{(1)}(x_i:\theta^*):\theta^*\};$$

it is *the Fisher information at* θ^* *for the multiple response* **x**.)

Now consider the analysis of a multiple-response vector $(x_1, \ldots, x_n)$. General familiarity with the application may suggest an initial reference value $\theta^{(0)}$. The limiting likelihood function appropriate to the reference value $\theta^{(0)}$ may indicate that the maximum likelihood value is elsewhere; the indicated position from the limiting likelihood is at the θ value given by

$$\theta - \theta^{(0)} = \frac{l^{(1)}(\mathbf{x}:\theta^{(0)})}{nj(\theta^{(0)})}.$$

Designate this value of θ by

$$\theta^{(1)} = \theta^{(0)} + \frac{l^{(1)}(\mathbf{x}:\theta^{(0)})}{nj(\theta^{(0)})}.$$

A similar analysis at the reference value $\theta^{(1)}$ may indicate that the maximum is again elsewhere:

$$\theta^{(2)} = \theta^{(1)} + \frac{l^{(1)}(\mathbf{x}:\theta^{(1)})}{nj(\theta^{(1)})}.$$

Typically, several iterations lead to a reference value θ^* located approximately at the maximum of the likelihood function. The approximating measurement model is then

$$\frac{e}{\sqrt{n}},$$

$$\frac{l^{(1)}(\mathbf{x}:\theta^*)}{nj(\theta^*)} = (\theta - \theta^*) + \frac{e}{\sqrt{n}}.$$

Analysis of the simple measurement model is given in Chapter One. Tests of significance concerning a θ value can be made by calculating the corresponding error value and comparing it with the error distribution. The structural distribution for the quantity θ is normal with variance $1/nj(\theta^*)$ and located at

$$\theta^* + \frac{l^{(1)}(\mathbf{x}:\theta^*)}{nj(\theta^*)}.$$

(If θ^* is the exact maximum, then $l^{(1)}(\mathbf{x}:\theta^*) = 0$ and the distribution is located at θ^*.)

The corresponding results for a vector quantity $\boldsymbol{\theta}$ can be stated briefly. Let $f(x:\boldsymbol{\theta})$ be a classical model satisfying the generalized Assumptions 1, 2, and 4. Consider a multiple response vector $(x_1, \ldots, x_n)$ with n large. For a value $\boldsymbol{\theta}^*$ near the actual $\boldsymbol{\theta}^0$ the large sample model can be approximated by the *simple measurement model:*

$$\mathbf{e},$$

$$\mathbf{w} = \boldsymbol{\tau} + \mathbf{e}.$$

The model has an error variable $\mathbf{e}$ *with a multivariate normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $J^{-1}(\boldsymbol{\theta}^*)$*; and it has a structural equation in which a realized value from the error distribution provides the link between the observed value of* $\mathbf{w}$,

$$\mathbf{w} = \sqrt{n}\,\frac{J^{-1}(\boldsymbol{\theta}^*)}{n}\,\mathbf{l}^D(\mathbf{x}:\boldsymbol{\theta}^*),$$

and the unknown value of

$$\boldsymbol{\tau} = \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}^*).$$

The measurement model can be expressed directly in terms of the quantity $\boldsymbol{\theta}$:

$$\frac{\mathbf{e}}{\sqrt{n}},$$

$$\frac{J^{-1}(\boldsymbol{\theta}^*)}{n}\,\mathbf{l}^D(\mathbf{x}:\boldsymbol{\theta}^*) = (\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \frac{\mathbf{e}}{\sqrt{n}}$$

(note that the matrix multiplying the likelihood gradient vector is the inverse of the Fisher information matrix,

$$nJ(\boldsymbol{\theta}^*) = E\{\mathbf{l}^D(\mathbf{x}:\boldsymbol{\theta}^*)\mathbf{l}'^D(\mathbf{x}:\boldsymbol{\theta}^*):\boldsymbol{\theta}^*\} = E\left\{\sum_{i=1}^{n} \mathbf{l}^D(x_i:\boldsymbol{\theta}^*)\mathbf{l}'^D(x_i:\boldsymbol{\theta}^*):\boldsymbol{\theta}^*\right\},$$

for the *multiple* response $\mathbf{x}$).

Now consider the analysis of a response vector $(x_1, \ldots, x_n)$. The approximating model at a reference value $\boldsymbol{\theta}^{(0)}$ may indicate that the maximum likelihood value is elsewhere:

$$\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \frac{J^{-1}(\boldsymbol{\theta}^{(0)})}{n}\,\mathbf{l}^D(\mathbf{x}:\boldsymbol{\theta}^{(0)}).$$

The model at $\boldsymbol{\theta}^{(1)}$ may indicate that the maximum is elsewhere:

$$\boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}^{(1)} + \frac{J^{-1}(\boldsymbol{\theta}^{(1)})}{n}\,\mathbf{l}^D(\mathbf{x}:\boldsymbol{\theta}^{(1)}).$$

Typically, a reference value $\boldsymbol{\theta}^*$ near the maximum point for the likelihood may be obtained in several iterations.

Tests of significance can be made by using the approximating measurement model. The structural distribution for $\boldsymbol{\theta}$ is multivariate normal with covariance matrix $(nJ(\boldsymbol{\theta}^*))^{-1}$ and located at

$$\boldsymbol{\theta}^* + (nJ(\boldsymbol{\theta}^*))^{-1}\mathbf{l}^D(\mathbf{x}:\boldsymbol{\theta}^*).$$

NOTES AND REFERENCES

The likelihood function was promoted and developed by R. A. Fisher; see Notes and References, Chapter Four. The proof in Section 1 that the likelihood, outside a neighborhood of the actual value of the quantity, goes to $-\infty$ with increasing sample size is derived from Wald (1949).

The limiting normality of the location of maximum likelihood was presented by Fisher (1922) and given heuristic proof in his subsequent papers. It has been widely examined in the literature.

The limiting quadratic form of the likelihood function has received general

recognition and is implicit in Fisher's treatment of likelihood, but it has apparently had no direct examination in the literature.

The strong law of large numbers and the convergence theorem for orthogonal variables may be found in Loève (1960).

The iterative procedure in Section 3 is based directly on the limiting likelihood function appropriate to the reference point being examined. It was proposed on other grounds by Fisher (e.g., 1956).

Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics, *Phil. Trans. Roy. Soc. London*, **A222**, 309–368; also as Paper 10 in Fisher (1950).
Fisher, R. A. (1950), *Contributions to Mathematical Statistics*, Wiley, New York.
Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, London.
Fraser, D. A. S. (1964a), Local conditional sufficiency, *J. Roy. Statist. Soc.*, **B26**, 52–62.
Fraser, D. A. S. (1964b), On local inference and information, *J. Roy. Statist. Soc.*, **B26**, 253–260.
Loève, M. (1960), *Probability Theory*, Van Nostrand, Princeton.
Wald, A. (1949), Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statistics*, **20**, 595–600.

PROBLEMS

***1.** Prove Lemma 8 in Section 2.2.

***2.** Suppose the classical model $f(x:\boldsymbol{\theta})$ satisfies the generalized Assumptions 2 and 4 in Section 2.2. In the pattern followed for a real quantity, derive the limiting form of the likelihood function for bounded $\boldsymbol{\tau}$,

$$\boldsymbol{\theta} = \boldsymbol{\theta}^0 + \boldsymbol{\tau} n^{-\frac{1}{2}},$$

where $\boldsymbol{\theta}^0$ designates the actual value of the quantity.

***3.** Consider the change from θ^0 as reference point to θ^* as reference point (Section 3). Let

$$\theta^* = \theta^0 + \delta n^{-\frac{1}{2}};$$

then

$$\tau^* = \tau - \delta.$$

Let

$$w^* = \frac{l^{(1)}(\mathbf{x}:\theta^*)}{\sqrt{nj(\theta^*)}}.$$

Show that

$$w^* = w - \delta + r_n,$$

where

$$\Pr\left\{\lim_{n\to\infty} r_n = 0 : \theta^0\right\} = 1.$$

Then show that the limiting form of the likelihood function is

$$-\frac{j(\theta^*)}{2}(w^* - \tau^*)^2 + \frac{j(\theta^*)}{2} w^{*2};$$

and show that w^* has a limiting normal distribution with location quantity τ^* and variance $1/j(\theta^*)$; the actual value for τ^* is $-\delta$.

4. Consider the simple measurement model

$$f(e)\, de,$$

$$x = \theta + e.$$

The corresponding classical model for the response variable x is

$$f(x - \theta)\, dx.$$

And the corresponding model for a multiple response is

$$\prod f(x_i - \theta) \prod dx_i.$$

Simplify Assumptions 1, 2, and 4 and express them in terms of properties of the error density.

5 (*Continuation*). Show that the likelihood function from $\mathbf{x}$ in the full model

$$R^+(\mathbf{x}) \prod f(x_i - \theta) \prod dx_i$$

is the same as the likelihood function from $\hat{\theta} = \hat{\theta}(\mathbf{x})$ in the conditional model.

6 (*Continuation*). Show that the location $\hat{\theta}(\mathbf{x})$ of maximum likelihood is a location variable. Show that the conditional model for $\hat{\theta}(\mathbf{x})$ given the orbit is

$$k(\mathbf{d}) \prod f(\hat{\theta} + d_i - \theta)\, d\hat{\theta}.$$

7 (*Continuation*). Under Assumptions 1, 2, and 4 show that with probability 1 the limiting *conditional* distribution of $\sqrt{n}\,(\hat{\theta}(\mathbf{x}) - \theta^0)$ is normal with mean 0 and variance $1/j(\theta^0)$. Thus the conditional analysis, given orbit, applied to the classical location model agrees in large samples with the likelihood analysis in Sections 2 and 3. (Fraser, 1964a,b.)

CHAPTER NINE

Precision and Information

The simple measurement model in Chapter One describes multiple measurements on a real-valued quantity θ; the error distribution of the measuring instrument is known.

If the error probability distribution in the reduced model is broad and diffuse, the measurements on θ can be called *imprecise*. Alternatively, if the error probability distribution is narrow and concentrated, the measurements can be called *precise*. This chapter examines the concept of *precision* for measurement models, for structural models, and for classical models with large-sample likelihood inference.

The simple measurement model produces a structural probability distribution for the quantity θ. A large value of the structural density function at a certain value for θ is information in favor of that value or *information for* that value. A small value of the structural density at a value for θ is *information against* that value. This chapter also examines the concept of *information* for measurement models and for structural models.

1 PRECISION: WITH A REAL-VALUED QUANTITY

Consider the simple measurement model but in a slightly generalized form. Let x_1 be a first measurement on θ with corresponding error distribution $f_1(e_1)\,de_1, \ldots$, and let x_n be an nth measurement on θ with corresponding error distribution $f_n(e_n)\,de_n$. The composite model,

$$\prod f_i(e_i) \prod de_i,$$
$$\mathbf{x} = [\theta, 1]\mathbf{e},$$

is a structural model.

Let $r(\mathbf{x})$ be a location variable and $d(\mathbf{x})$ be the reference point; the

reduced model is

$$k(\mathbf{d}(\mathbf{x})) \prod f_i(r + d_i(\mathbf{x}))\,dr,$$
$$r(\mathbf{x}) = \theta + r.$$

The case of *normal* error again leads to special simplicities: the conditional error distribution is the *same on each orbit*; and the conditional error distribution is *normal* but more concentrated. Let the ith error variable be normal with known variance σ_i^2:

$$f_i(e_i)\,de_i = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left\{-\frac{1}{2}\frac{e_i^2}{\sigma_i^2}\right\} de_i.$$

The conditional error distribution is

$$g(r:\mathbf{d})\,dr = k' \frac{1}{(2\pi)^{n/2} \prod \sigma_i} \exp\left\{-\frac{1}{2}\sum\left(\frac{r + d_i}{\sigma_i}\right)^2\right\} dr$$
$$= k'' \exp\left\{-\tfrac{1}{2}r^2 \sum\left(\frac{1}{\sigma_i^2}\right) - r\sum\left(\frac{d_i}{\sigma_i^2}\right)\right\} dr.$$

This distribution for the *reduced* model has normal form; its variance is σ_R^2, where

$$\frac{1}{\sigma_R^2} = \sum \frac{1}{\sigma_i^2};$$

but its location depends on the choice of reference point. A convenient reference point is the center of the conditional distribution; this requires

$$\sum \frac{d_i}{\sigma_i^2} = 0$$

and leads to

$$r(\mathbf{x}) = \frac{\sum (x_i/\sigma_i^2)}{\sum (1/\sigma_i^2)} = \sum x_i \frac{\sigma_R^2}{\sigma_i^2}$$

(the coordinates are *weighted in proportion to reciprocal variances*). With this choice for the location variable, the conditional distribution becomes

$$g(r:\mathbf{d})\,dr = k \exp\left\{-\frac{1}{2}\frac{r^2}{\sigma_R^2}\right\} dr$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma_R} \exp\left\{-\frac{1}{2}\frac{r^2}{\sigma_R^2}\right\} dr;$$

this distribution is the *same on each orbit*; and it is *normal* with mean zero and variance σ_R^2,

$$\frac{1}{\sigma_R^2} = \sum_1^n \frac{1}{\sigma_i^2}.$$

The reduced measurement model is

$$\sum x_i \frac{\sigma_R^2}{\sigma_i^2} \overset{r,}{=} \theta + r,$$

where r designates a normal variable with mean 0 and variance σ_R^2.

Thus, with normal error components, the error variable in the reduced model is also normal; and the reciprocal variance for the error in the reduced model is the sum of the reciprocal variances for the components:

$$\frac{1}{\sigma_R^2} = \sum_1^n \frac{1}{\sigma_i^2}.$$

A small reciprocal variance gives a diffuse distribution and implies imprecise measurement. A large reciprocal variance gives a concentrated distribution and implies precise measurement.

For a normal error distribution the precision is defined to be the reciprocal variance.

Let j_R be the precision in the reduced model and j_i be the precision of the ith error component:

$$j_R = \frac{1}{\sigma_R^2}, \qquad j_i = \frac{1}{\sigma_i^2}.$$

Then, *with normal error components in the measurement model, the error in the reduced model is also normal and its precision is the sum of the component precisions:*

$$j_R = j_1 + \cdots + j_n.$$

Now consider a sequence of response variables $x_1, \ldots, x_n$ with corresponding classical models $f_1(x_1:\theta), \ldots, f_n(x_n:\theta)$ involving a real quantity θ. Suppose that some generalized assumptions are fulfilled that cover the extension† of the limiting-likelihood results to variables with differing distributions. Then, as the number of response variables approaches infinity, the likelihood function approaches normal quadratic form and the model admits approximation by the simple measurement model

$$\frac{\sum l_i^{(1)}(x_i:\theta^*)}{\sum j_i(\theta^*)} \overset{r,}{=} (\theta - \theta^*) + r,$$

† The central limit theorem and the law of large numbers have extensions for differently distributed variables; the details for the generalized assumptions are not of importance here.

where r designates a normal variable with mean 0 and *precision* $\sum j_i(\theta^*)$, where

$$l_i^{(1)}(x_i:\theta^*) = \frac{\partial}{\partial\theta} l_i(x_i:\theta^*) = \frac{\partial}{\partial\theta} \ln f_i(x_i:\theta^*),$$

$$\begin{aligned} j_i(\theta^*) &= \text{var}\,\{l_i^{(1)}(x_i:\theta^*):\theta^*\} \\ &= E\{-l_i^{(11)}(x_i:\theta^*):\theta^*\}, \end{aligned}$$

and where θ^* is in a small neighborhood of the actual θ^0.

Thus *the error in the approximating model is normal and its precision $j_R(\theta^*)$ is the sum of components $j_1(\theta^*), \ldots, j_n(\theta^*)$, one from each of the component variables.* The approximating model is as if each component variable x_i were a measurement variable with normal error and with precision $j_i(\theta^*)$.

For a classical model $f(x:\theta)$ satisfying Assumptions 1, 2, and 4 *in Chapter Eight, the precision at the value θ* is defined to be *the variance of the likelihood derivative:*

$$j(\theta) = \text{var}\,\{l^{(1)}(x:\theta):\theta\}.$$

When independent variables involving the same quantity θ are combined, the precisions are added to obtain the precision of the composite variable. The precision for a large number of variables is equal to the precision of the approximating normal measurement model. For a small number of variables, specifically measurement variables, the conditional error distribution depends typically on the orbit; the concept of precision is inadequate. A concept of information in Section 3 is then needed to give a general description of the measurement process.

2 PRECISION: WITH A VECTOR-VALUED QUANTITY

Consider the simple measurement model extended to cover a vector quantity $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$. For an ith measurement let

$$f_i(e_{1i}, \ldots, e_{ki})\, de_{1i} \cdots de_{ki} = f_i(\mathbf{e}_i)\, d\mathbf{e}_i$$

be the error distribution, and let $\mathbf{x}_i = \boldsymbol{\theta} + \mathbf{e}_i$,

$$\begin{bmatrix} x_{1i} \\ \cdot \\ \cdot \\ \cdot \\ x_{ki} \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_k \end{bmatrix} + \begin{bmatrix} e_{1i} \\ \cdot \\ \cdot \\ \cdot \\ e_{ki} \end{bmatrix},$$

be the structural equation; the group is the location group on R^k:

$$G = \{[\mathbf{a}, I]:\quad -\infty < a_j < \infty\},$$

where

$$[\mathbf{a}, I]\mathbf{x} = \mathbf{a} + \mathbf{x}.$$

The multiple model for n measurements is

$$\prod f_i(\mathbf{e}_i) \prod d\mathbf{e}_i,$$
$$(\mathbf{x}_1, \ldots, \mathbf{x}_n) = [\boldsymbol{\theta}, I](\mathbf{e}_1, \ldots, \mathbf{e}_n);$$

this is a slightly generalized form of the *simple composite-measurement model* in Problem 8, Chapter Four.

Let $\mathbf{r}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be a location variable:

$$r(\mathbf{a} + \mathbf{x}_1, \ldots, \mathbf{a} + \mathbf{x}_n) = \mathbf{a} + \mathbf{r}(\mathbf{x}_1, \ldots, \mathbf{x}_n),$$

and let $\mathbf{d}_1, \ldots, \mathbf{d}_n$ be the corresponding deviations:

$$\mathbf{d}_i(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathbf{x}_i - \mathbf{r}(\mathbf{x}_1, \ldots, \mathbf{x}_n).$$

The reduced model is

$$k(\mathbf{d}_1, \ldots, \mathbf{d}_n) \prod f_i(\mathbf{r} + \mathbf{d}_i)\, d\mathbf{r},$$
$$\mathbf{r}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \boldsymbol{\theta} + \mathbf{r}.$$

The X, E notation would be simpler, but it is more convenient here to have separate designations for the individual measurements.

Consider the case of normal error. Let $\mathbf{e}_i$ be multivariate normal with known inverse covariance matrix J_i:

$$f_i(\mathbf{e}_i)\, d\mathbf{e}_i = \frac{|J_i|^{1/2}}{(2\pi)^{k/2}} \exp\{-\tfrac{1}{2}\mathbf{e}_i' J_i \mathbf{e}_i\}\, d\mathbf{e}_i.$$

The conditional error distribution is

$$\begin{aligned} g(\mathbf{r}:\mathbf{d}_1, \ldots, \mathbf{d}_n)\, d\mathbf{r} &= k' \exp\{-\tfrac{1}{2} \textstyle\sum (\mathbf{r} + \mathbf{d}_i)' J_i (\mathbf{r} + \mathbf{d}_i)\}\, d\mathbf{r} \\ &= k'' \exp\left\{-\tfrac{1}{2}\mathbf{r}' \sum_1^n J_i \mathbf{r} - \mathbf{r}' \sum_1^n J_i\, \mathbf{d}_i\right\} d\mathbf{r}. \end{aligned}$$

This reduced-model distribution is also multivariate normal; its inverse covariance matrix is

$$J_R = \sum_1^n J_i,$$

but its location depends on the choice of reference point. A convenient reference point is the center of the conditional distribution; this requires

$$\sum J_i \mathbf{d}_i = 0$$

and leads to

$$\mathbf{r}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = J_R^{-1} \sum_1^n J_i \mathbf{x}_i$$

(the coordinates are *weighted in proportion to inverse covariance matrices*). With this choice for the location variable, the conditional distribution becomes

$$g(\mathbf{r}:\mathbf{d}_1, \ldots, \mathbf{d}_n)\, d\mathbf{r} = \frac{|J_R|^{1/2}}{(2\pi)^{k/2}} \exp\{-\tfrac{1}{2}\mathbf{r}' J_R \mathbf{r}\}\, d\mathbf{r};$$

this distribution is *the same on each orbit*; and it is *multivariate normal* with mean 0 and inverse covariance matrix

$$J_R = \sum_1^n J_i.$$

The reduced measurement model is

$$\mathbf{r},$$
$$J_R^{-1} \sum_1^n J_i \mathbf{x}_i = \boldsymbol{\theta} + \mathbf{r},$$

where the error $\mathbf{r}$ is multivariate normal with inverse covariance matrix $J_R = \Sigma J_i$.

For a normal error distribution the precision is defined to be the inverse covariance matrix.

The precision for the ith component is J_i; the precision for the reduced model is J_R:

$$J_R = \sum_1^n J_i.$$

Thus, *with normal error components, the reduced model has normal error and its precision is the sum of the component precisions.*

Now consider a sequence of response variables $x_1, \ldots, x_n$ with corresponding classical models $f_1(x_1:\boldsymbol{\theta}), \ldots, f_n(x_n:\boldsymbol{\theta})$ involving the vector quantity $\boldsymbol{\theta}$. Suppose that generalized assumptions are fulfilled that ensure the limiting-normal likelihood function and the approximation by the simple measurement model:

$$\mathbf{r},$$
$$(\textstyle\sum J_i(\boldsymbol{\theta}^*))^{-1} \sum \mathbf{l}_i^D(x_i:\boldsymbol{\theta}^*) = (\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \mathbf{r},$$

where $\mathbf{r}$ designates a multivariate normal variable with mean $\mathbf{0}$ and precision $J_R(\boldsymbol{\theta}^*) = \Sigma J_i(\boldsymbol{\theta}^*)$, where

$$\mathbf{l}_i^D(x_i:\boldsymbol{\theta}^*) = \frac{\partial}{\partial \boldsymbol{\theta}} l_i(x_i:\boldsymbol{\theta}^*) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_i(x_i:\boldsymbol{\theta}^*),$$
$$J_i(\boldsymbol{\theta}^*) = E\{\mathbf{l}^D(x_i:\boldsymbol{\theta}^*)\mathbf{l}'^D(x_i:\boldsymbol{\theta}^*):\boldsymbol{\theta}^*\},$$

and where $\boldsymbol{\theta}^*$ is in a small neighborhood of the actual $\boldsymbol{\theta}^0$.

Thus *the error in the approximating model is normal, and its precision* $J_R(\boldsymbol{\theta}^*)$ *is the sum of the components* $J_1(\boldsymbol{\theta}^*), \ldots, J_n(\boldsymbol{\theta}^*)$. The approximating model is as if each component variable x_i were a measurement variable with normal error and precision $J_i(\boldsymbol{\theta}^*)$.

For a classical model satisfying generalized Assumptions 1, 2, and 4 *in Chapter Eight, the precision matrix at the value* $\boldsymbol{\theta}$ *is*

$$J(\boldsymbol{\theta}) = E\{\mathbf{l}^D(x:\boldsymbol{\theta})\mathbf{l}'^D(x:\boldsymbol{\theta}):\boldsymbol{\theta}\};$$

an alternative expression involving second derivatives is given in Section 2, Chapter Eight.

When independent variables involving a quantity $\boldsymbol{\theta}$ are combined, the precision matrices are added to obtain the precision matrix for the composite variable. For a large number of variables the precision is the precision of the approximating normal measurement model.

3 INFORMATION: THE SIMPLE MEASUREMENT MODEL

Consider the simple measurement model

$$\prod f(e_i) \prod de_i,$$
$$\mathbf{x} = \theta\mathbf{1} + \mathbf{e},$$

and the corresponding reduced model

$$k(\mathbf{d}(\mathbf{x})) \prod f(r + d_i(\mathbf{x}))\, dr,$$
$$r(\mathbf{x}) = \theta + r.$$

With various error distributions and with various observed orbits, a broad range of conditional error distributions is possible. The concept of *precision* is useful with normal errors or with a large number of error components; for other cases the more general concept of *information* is needed to effectively assess measurements and the measurement process.

The structural probability element for θ given the measurement vector $\mathbf{x}$ is

$$g^*(\theta:\mathbf{x})\, d\theta = k(\mathbf{d}(\mathbf{x})) \prod f(x_i - \theta)\, d\theta;$$

this probability element gives the probability that the quantity is in a neighborhood $d\theta$ of θ. The level of this probability can be described conveniently in logarithmic units: *the information for the value* θ *given the measurement vector* $\mathbf{x}$ is

$$I(\mathbf{x}, \theta) = \ln g^*(\theta:\mathbf{x}) = \ln\left(k(\mathbf{d}(\mathbf{x})) \prod f(x_i - \theta)\right).$$

A large positive value for the information $I(\mathbf{x}, \theta)$ is information in favor of the θ value; a large negative value is information against; the value 0 corresponds to unit structural density.

Note that the difference in information between two θ-values is equal to the log-likelihood difference:

$$\begin{aligned} I(\mathbf{x}, \theta'') - I(\mathbf{x}, \theta') &= \ln\left(k(\mathbf{d}(\mathbf{x})) \prod f(x_i - \theta'')\right) - \ln\left(k(\mathbf{d}(\mathbf{x})) \prod f(x_i - \theta')\right) \\ &= \ln\left(\prod f(x_i - \theta'')\right) - \ln\left(\prod f(x_i - \theta')\right) \\ &= l(\mathbf{x}:\theta'') - l(\mathbf{x}:\theta'). \end{aligned}$$

The information function is thus a representative log-likelihood function, but it has vertical placement—*it has a zero point on the vertical scale.*

For the simple measurement model having normal error with variance σ_0^2 the conditional error distribution is

$$g(\bar{e}:\mathbf{d})\, d\bar{e} = \left(\frac{1}{2\pi}\frac{n}{\sigma_0^2}\right)^{1/2} \exp\left\{-\tfrac{1}{2}\bar{e}^2\frac{n}{\sigma_0^2}\right\} d\bar{e}.$$

The information for θ given $\mathbf{x}$ is then

$$I(\mathbf{x}, \theta) = \tfrac{1}{2}\ln\left(\frac{n}{\sigma_0^2}\right) - \tfrac{1}{2}\ln(2\pi) - \tfrac{1}{2}(\bar{x} - \theta)^2\left(\frac{n}{\sigma_0^2}\right)$$

(see Figure 1).

The information $I(\mathbf{x}, \theta)$ describes *information for* θ *given the measurement vector* $\mathbf{x}$. Consider now the measurement *process* and its overall effectiveness. Let θ^0 designate the actual value of the quantity. Then, for the process of making n measurements on the quantity θ^0, let

$$I(\theta^0; \theta) = E\{I(\mathbf{x}, \theta):\theta^0\}$$

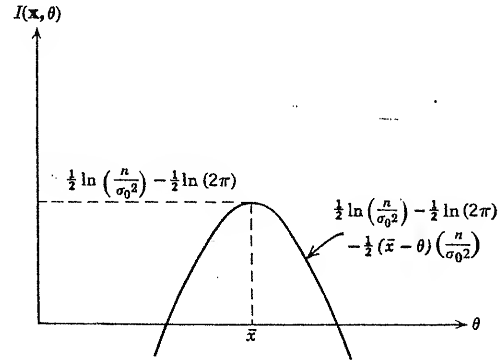


Figure 1. The information $I(\mathbf{x}, \theta)$ for θ, given $\mathbf{x}$ in the case of normal error.

be *the information for the value θ given the quantity θ^0*; it is the mean value of the information $I(\mathbf{x}, \theta)$ at θ when the quantity has value θ^0.

$$I(\theta^0; \theta) = \int \ln g^*(\theta : \mathbf{x}) \prod f(x_i - \theta^0) \prod dx_i$$

$$= \int \ln \left(k(\mathbf{d}(\mathbf{x})) \prod f(x_i - \theta)\right) \prod f(x_i - \theta^0) \prod dx_i$$

$$= \int \ln \left(k(\mathbf{d}(\mathbf{x})) \prod f(e_i - \delta)\right) \prod f(e_i) \prod de_i.$$

Note that the information at θ given θ^0 depends only on the difference

$$\delta = \theta - \theta^0.$$

The information for the actual value θ^0 is perhaps the most significant indicator of the effectiveness of the measurement process: $I(\theta^0; \theta^0)$.

For the simple measurement model with normal error the information at θ given θ^0 is

$$I(\theta^0; \theta) = \tfrac{1}{2} \ln \left(\frac{n}{\sigma_0^2}\right) - \tfrac{1}{2} \ln (2\pi) - \frac{1}{2}\left(\frac{n}{\sigma_0^2}\right) E\{(\bar{e} - \delta)^2\}$$

$$= \tfrac{1}{2} \ln \left(\frac{n}{\sigma_0^2}\right) - \tfrac{1}{2} \ln (2\pi) - \frac{1}{2}\left(1 + \delta^2 \frac{n}{\sigma_0^2}\right)$$

(see Figure 2). The information for the actual value is

$$I(\theta^0; \theta^0) = \tfrac{1}{2} \ln \left(\frac{n}{\sigma_0^2}\right) - \tfrac{1}{2} \ln (2\pi) - \tfrac{1}{2};$$

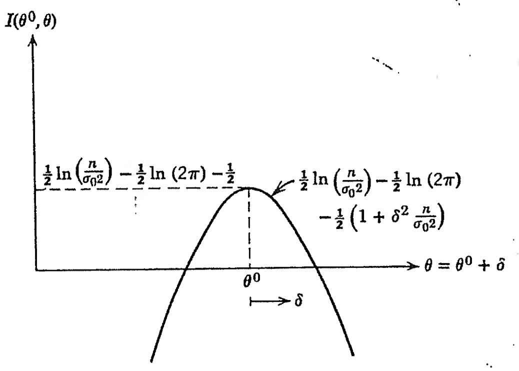


Figure 2. The information at $\theta = \theta^0 + \delta$, given the actual value θ^0.

it is an increasing function of the precision n/σ_0^2. Thus with normal error, the *information for the actual value* and the *precision* are one-to-one monotone functions each of the other.

Consider some properties of the information $I(\theta^0; \theta)$.

The information has a maximum value at the actual θ^0:

$$I(\theta^0; \theta) - I(\theta^0; \theta^0) = E\{I(\mathbf{x}, \theta) - I(\mathbf{x}, \theta^0) : \theta^0\}$$

$$= E\{l(\mathbf{x} : \theta) - l(\mathbf{x} : \theta^0) : \theta^0\} < 0;$$

the inequality follows from Lemma 1 in Chapter Eight.

The curvature of the information at the actual value is equal to the *precision*: Suppose Assumptions 1, 2, and 4 in Chapter Eight hold; then

$$\left[\frac{\partial}{\partial \theta} I(\theta^0; \theta)\right]_{\theta=\theta^0} = 0,$$

$$-\frac{\partial^2}{\partial \theta^2} I(\theta^0; \theta) = -E\left\{\frac{\partial^2}{\partial \theta^2} \ln\left(k(\mathbf{d}(\mathbf{x})) \prod f(x_i - \theta)\right) : \theta^0\right\} = -E\left\{\frac{\partial^2}{\partial \theta^2} l(\mathbf{x} : \theta) : \theta^0\right\},$$

$$\left[-\frac{\partial^2}{\partial \theta^2} I(\theta^0; \theta)\right]_{\theta=\theta^0} = nj(\theta^0).$$

*4 INFORMATION: THE STRUCTURAL MODEL

Consider the structural model

$$f(E)\, dE,$$

$$X = \theta E,$$

and the corresponding reduced model

$$g([E] : D(X))\, d[E],$$

$$[X] = \theta[E];$$

the quantity θ is an element of a group G (Assumption 3, Chapter 2), and the conditional error distribution is

$$g([E] : D)\, d[E] = k(D) f([E]D) \frac{J_N(E)}{J_L([E])}\, d[E].$$

The structural probability element for the quantity θ is

$$g^*(\theta : X)\, d\theta = k(D(X)) f(\theta^{-1}X) J_N(\theta^{-1}X) \frac{J_L^*([X])}{J_L([X])} \frac{d\theta}{J_L^*(\theta)}$$

$$= k(D(X)) f(\theta^{-1}X) J_N(\theta^{-1}X) \frac{J_L^*([X])}{J_L([X])}\, d\nu(\theta);$$

the right invariant differential $d\nu(\theta)$ for θ corresponds to the left invariant differential $d\mu([E])$ for $[E]$. The density function with respect to $d\nu(\theta)$ can be described conveniently in logarithmic units: *the information for the value θ given the observation X is*

$$I(X, \theta) = \ln\left(k(D(X))f(\theta^1X)J_N(\theta^{-1}X)\frac{J_L^*([X])}{J_L([X])}\right).$$

A large positive value is information *for* the θ value; a large negative value is information *against* the θ value.

The difference in information between two θ values is equal to the likelihood difference for the two θ values: the classical model for the response X is

$$f(\theta^{-1}X)J_N(\theta^{-1}X)\frac{dX}{J_N(X)};$$

the log-likelihood function is

$$l(X:\theta) = R(X) + \ln\left(f(\theta^{-1}X)J_N(\theta^{-1}[X])\right);$$

the information difference from θ' to θ'' is

$$\begin{aligned} I(X, \theta'') - I(X, \theta') &= \ln\left(f(\theta''^{-1}X)J_N(\theta''^{-1}[X])\right) - \ln\left(f(\theta'^{-1}X)J_N(\theta'^{-1}[X])\right) \\ &= l(X:\theta'') - l(X:\theta'), \end{aligned}$$

which is the likelihood difference from θ' to θ''.

The information $I(X, \theta)$ describes information for θ *given* X. Let θ^0 denote the actual value of the quantity. Then *the information for θ given θ^0* is defined as

$$\begin{aligned} I(\theta^0; \theta) &= E\{I(X, \theta):\theta^0\} \\ &= \int \ln\left(k(D(X))f(\theta^{-1}X)J_N(\theta^{-1}X)\frac{J_L^*([X])}{J_L([X])}\right)\cdot f(\theta^{0-1}X)J_N(\theta^{0-1})\, dX. \end{aligned}$$

The information for θ has a maximum value at the actual θ^0:

$$\begin{aligned} I(\theta^0; \theta) - I(\theta^0; \theta^0) &= E\{l(X:\theta) - l(X:\theta^0):\theta^0\} \\ &= E\left\{\ln\frac{f(\theta^{-1}X)J_N(\theta^{-1})}{f(\theta^{0-1}X)J_N(\theta^{0-1})}:\theta^0\right\} < 0; \end{aligned}$$

the inequality follows from Lemma 1 in Chapter Eight.

The curvature matrix of the information at the actual value is equal to the precision matrix. Suppose the generalized Assumptions 1, 2, and 4 in Chapter Eight hold; then the jj' element in the curvature matrix is

$$\begin{aligned} \left[-\frac{\partial^2}{\partial\theta_j\,\partial\theta_{j'}}I(\boldsymbol{\theta}^0;\boldsymbol{\theta})\right]_{\theta=\theta^0} &= -E\left\{\frac{\partial^2}{\partial\theta_j\,\partial\theta_{j'}}l(X:\boldsymbol{\theta}^0):\boldsymbol{\theta}^0\right\} \\ &= -E\{l^{(jj')}(X:\boldsymbol{\theta}^0):\boldsymbol{\theta}^0\}, \end{aligned}$$

which is the jj' element in the precision matrix.

NOTES AND REFERENCES

The concept of precision was introduced as intrinsic accuracy by Fisher (1922). The concept of information for measurement and structural model was introduced by Fraser (1965); for these models it embraces three earlier concepts of information: Fisher (1922, 1956), Shannon (1948), and Kullback (1959). The Fisher information function is defined in Chapter Eight. The Shannon information function $I_x^S(\theta)$ for a variable x with model $f(x:\theta)$ is

$$\int \ln f(x:\theta)f(x:\theta)\,dx.$$

The Kullback information $I^K(\theta^0, \theta)$ is a mean value of a log-likelihood difference:

$$I^K(\theta^0; \theta) = \int\left(\ln f(x:\theta^0) - \ln f(x:\theta)\right)f(x:\theta^0)\,dx.$$

Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics, *Phil. Trans. Roy. Soc. London*, **A222,** 309–368; also as Paper 10, Fisher (1950).
Fisher, R. A. (1950), *Contributions to Mathematical Statistics*, Wiley, New York.
Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, London.
Fraser, D. A. S. (1965), On information in statistics, *Ann. Math. Statistics*, **36,** 890–896.
Kullback, S. (1959), *Information Theory and Statistics, Wiley, New York.*
Shannon, C. E. (1948), A mathematical theory of communication, *Bell System Tech. J.*, **27,** 379–423, 623–656.

PROBLEMS

1. Show that the precision of the Cauchy error distribution

$$\frac{1}{\pi}\frac{de}{1+e^2}, \quad -\infty < e < \infty,$$

is equal to the precision of the normal error distribution with mean 0 and variance 2. Use the integral $\int_{-\infty}^{\infty}(1+t^2)^{-2}\,dt = \pi/2$. (Fisher, 1922.)

2. Consider the simple measurement model with uniform error:

$$\begin{aligned} f(e) &= 1, \quad && -\tfrac{1}{2} < e < \tfrac{1}{2}, \\ &= 0, && \text{otherwise.} \end{aligned}$$

(i) Show that

$$\begin{aligned} I(x, \theta) &= -\ln(1 - R), \quad && |r(x) - \theta| \le \frac{1-R}{2}, \\ &= -\infty, && \text{otherwise,} \end{aligned}$$

where the location variable is

$$r(\mathbf{x}) = \frac{\max x_i + \min x_i}{2}$$

and $R = \max x_i - \min x_i$.

(ii) Show that

$$I(\theta^0; \theta) = -\infty, \qquad \text{if } \theta \neq \theta^0.$$

(iii) For one measurement on θ show that

$$I(\theta^0; \theta^0) = 0;$$

and for two measurements on θ show that

$$I(\theta^0; \theta^0) = \tfrac{1}{2}.$$

3. Consider the simple measurement model with location variable $r(\mathbf{x}) = x_1$ and

$$\mathbf{d}(\mathbf{x}) = (0, d_2(\mathbf{x}), \ldots, d_n(\mathbf{x})) = (0, x_2 - x_1, \ldots, x_n - x_1).$$

The error distribution in terms of $(e_1, \ldots, e_n)$ can be expressed in terms of $e_1, d_2, \ldots, d_n$:

$$\prod f(e_i) = g(e_1 : \mathbf{d}) h(d_2, \ldots, d_n).$$

Show that

$$I(\theta^0; \theta^0) = I^S_{\mathbf{x}} - I^S_{\mathbf{d}};$$

the information at the actual value is the Shannon information in the original error distribution less the Shannon information in the orbital variable (whose value can be observed).

Answers to Selected Problems

CHAPTER ONE

1. (i) Reduced model: $\bar{e} = 0.7906z$; $61.5 = \theta + \bar{e}$.
(ii) $\Pr\{-1.55 \le \bar{e} \le 1.55\} = 95\%$;
$\Pr\{-2.04 \le \bar{e} \le 2.04\} = 99\%$.
(iii) The quantity θ is normal with mean 61.5 and standard deviation 0.7906.

$$\Pr\{59.95 \le \theta \le 63.05\} = 95\%; \; \Pr\{59.46 \le \theta \le 63.54\} = 99\%.$$

3. (i) Reduced model: $g(e_1)\,de_1$; $157.01 = \theta + e_1$, where $g(e_1)\,de_1 = 50de_1$ on $(-0.5, -0.48)$ and $= 0$ otherwise. Structural distribution: uniform on the interval (157.49, 157.51).
(ii) Reduced model and structural distribution as in (i); no.
(iii) The hypothesis $\theta = 157.60$ leads to $e_1 = -0.59$, a value outside the range of the error probability distribution; within the model the hypothesis is effectively denied.

5. (i) Reduced model: $g(r)\,dr$; $r(\mathbf{x}) = \theta + r$, where $g(r)\,dr = n\exp\{-nr\}\,dr$ on $(0, \infty)$ and $= 0$ otherwise.
(ii) The structural distribution is $n\exp\{-nr(\mathbf{x}) + n\theta\}\,d\theta$ on $(-\infty, r(\mathbf{x}))$ and 0 elsewhere.

9. The hypothesis leads to the value 2.4 for $\chi_3/\sqrt{3}$ or the value 17.28 for χ_3^2. The 1 and $\frac{1}{2}\%$ points are 11.3 and 12.8; within the model there is strong evidence that the hypothesis is not true.

12. The reference point has one $d = 0$, the remaining d's greater than 0, and the sum of the remaining d's equal to $n - 1$.

(i) Reduced model: $g(b, s)\,db\,ds$; $[b(\mathbf{x}), s(\mathbf{x})] = [\mu, \sigma][b, s]$, where

$$g(b, s) = n\exp\{-nb\}\frac{(n-1)^{n-1}}{\Gamma(n-1)}\exp\{-(n-1)s\}s^{n-2}\begin{cases}0 < b < \infty,\\ 0 < s < \infty.\end{cases}$$

(ii) The distribution of $t = b(\mathbf{e})/s(\mathbf{e})$ is

$$g_L(t) = n\left(1 + \frac{n}{n-1}t\right)^{-n}, \qquad 0 < t < \infty.$$

The structural distribution for μ is

$$g_L\left(\frac{b(\mathbf{x}) - \mu}{s(\mathbf{x})}\right)\frac{d\mu}{s(\mathbf{x})}, \qquad -\infty < \mu < x_{(1)}.$$

15. (i) Reduced model: $g(b, s)\,db\,ds$; $[x_{(1)}, s(\mathbf{x})] = [\mu, \sigma][b, s]$, where

$$g(b, s) = k(\mathbf{d})\beta^n \prod (b + sd_i)^{\beta-1} \exp\{-\sum (b + sd_i)^\beta\} s^{n-2}$$

on $0 < b < \infty$, $0 < s < \infty$, and $= 0$ otherwise. The normalizing constant (partly evaluated) is

$$k(\mathbf{d}) = \left(\Gamma(n)\beta^{n-1} \int_0^\infty \frac{\Pi(t + d_i)^{\beta-1}}{(\sum (t + d_i)^\beta)^n}\,dt\right)^{-1}.$$

(ii) The structural distribution for $[\mu, \sigma]$ is

$$g\left(\frac{x_{(1)} - \mu}{\sigma}, \frac{s(\mathbf{x})}{\sigma}\right) \frac{s(\mathbf{x})}{\sigma^3}\,d\mu\,d\sigma,$$

where $g(b, s)$ is given in (i).

22. (i) The orbits are rays from the origin (origin excluded). The reference points are the points of intersection of orbits with the plane $\Sigma x_i = n$.

(ii) The error probability distribution is $\Gamma^{-1}(n)n^n \exp\{-n\bar{e}\}\bar{e}^{n-1}\,d\bar{e}$ on $(0, \infty)$.

(iii) The structural distribution for θ is

$$\frac{n^n}{\Gamma(n)} \exp\left\{-\frac{\sum x_i}{\theta}\right\}\left(\frac{\sum x_i}{n\theta}\right)^n \frac{d\theta}{\theta}.$$

CHAPTER TWO

12. (i) Suppose $g \in g_1H, g_2H$. Then $g = g_1h_1 = g_2h_2$ with $h_1, h_2 \in H$; $g_1H = gh_1^{-1}H = gH = gh_2^{-1}H = g_2H$. Otherwise, g_1H and g_2H are disjoint. It follows that $\{gH : g \in G\}$ is a partition of G.

(ii) The sets gH, Hg each contain g. Left partition is equal to right partition $\leftrightarrow gH = Hg$ for all $g \leftrightarrow H$ is normal.

13. Let $H = H_\alpha$, where $i \in H_\alpha$. For h in H: hH contains h; $hH \in \{H_\alpha\}$; hence $hH = H$. For h in H: $hH = H$ contains i; $h^{-1} \in H$. Thus H contains i and is closed under formation of products and inverses: H is a subgroup. It follows that $\{H_\alpha\}$ consists of left cosets of H.

CHAPTER THREE

10. Let $g = \begin{bmatrix} I & 0 \\ \mathbf{a}' & c \end{bmatrix}$ be a general element in G:

$$gL = g\left\{\begin{bmatrix} I & 0 \\ \mathbf{b}' & 1 \end{bmatrix} : \mathbf{b} \in R^r\right\} = \left\{\begin{bmatrix} I & 0 \\ \mathbf{a}' + c\mathbf{b}' & c \end{bmatrix} : \mathbf{b} \in R^r\right\} = \left\{\begin{bmatrix} I & 0 \\ \mathbf{b}' & c \end{bmatrix} : \mathbf{b} \in R^r\right\};$$

$$Lg = \left\{\begin{bmatrix} I & 0 \\ \mathbf{b}' & 1 \end{bmatrix} : \mathbf{b} \in R^r\right\} g = \left\{\begin{bmatrix} I & 0 \\ \mathbf{b}' + \mathbf{a}' & c \end{bmatrix} : \mathbf{b} \in R^r\right\} = \left\{\begin{bmatrix} I & 0 \\ \mathbf{b}' & c \end{bmatrix} : \mathbf{b} \in R^r\right\}.$$

Thus $gL = Lg$ for all g and L is normal.

12. The analysis-of-variance table is

Source	Dimension	Component	Structure of Component
Mean	1	814.088000	$(\alpha_1\sqrt{5} + \sigma z_1)^2$
Linear	1	1.102811	$(\alpha_2\sqrt{0.148} + \sigma z_2)^2$
Quadratic	1	0.003681	$(\alpha_3\sqrt{0.003670} + \sigma z_3)^2$
Residual	2	0.005508	$(\sigma\chi_2)^2$
Total	5	815.200000	

(i) On hypothesis $\beta_3(= \alpha_3) = 0$, $F = 0.003681/(0.005508/2) = 1.337$ is an observed value of F on 1 over 2, a reasonable value; data in accord with hypothesis.

(ii) On assumption $\alpha_3 = 0$, the quantity α_2 is now $\beta_2^{(2)}$ based on two structural vectors; the fitted general level is $5.06216 + 2.72973x = 12.76 + 2.72973(x - 2.82)$; the structural distribution is

$$\beta_2 = 2.72973 - \frac{\sqrt{0.005508}}{\chi_2} \frac{z_2}{\sqrt{0.148}} = 2.730 + t_2^*(0.136),$$

where t_2^* is ordinary t on 2 (precautionary analysis using residual length from three structural vectors).

17. (i) The convenient transformation variable and reference point are

$$[Y] = \begin{bmatrix} I & 0 \\ \mathbf{b}'(\mathbf{y}) & 1 \end{bmatrix}, \quad D(Y) = \begin{bmatrix} V \\ \mathbf{d}'(\mathbf{y}) \end{bmatrix},$$

where $\mathbf{b}(\mathbf{y})$ is the vector of regression coefficients of $\mathbf{y}$ on the row vectors in V and $\mathbf{d}(\mathbf{y})$ is the unadjusted residual vector.

(ii) $dm(Y) = \Pi\,dy_i$; $\quad d\mu(g) = d\nu(g) = \Pi\,db_u$; $\quad \Delta(g) \equiv 1$.

(iii) Error: $k(\mathbf{d})\Pi f(\Sigma\, b_u v_{ui} + d_i)\Pi\,db_u$. Structural: $k(\mathbf{d})\Pi f(y_i - \Sigma\beta_u v_{ui})\Pi\,d\beta_u$.

20. (i) The analysis of variance table is

Source	Dimension	Component
Mean	1	32,961.63
A	2	22.38
Residual	10	16.99
	13	33,001.00

(ii) The structural distributions are described by

$$\mu_3 - \frac{\mu_1 + \mu_2}{2} = 2.22 + 0.7431t_{10}^*,$$

$$\mu_2 - \mu_1 = 1.90 + 0.9217t_{10}^{**},$$

where t_{10}^*, t_{10}^{**} are ordinary t variables on 10 degrees of freedom.

26. The marginal structural distributions are

$$g_L^*(\boldsymbol{\mu} : Y)\,d\boldsymbol{\mu} = k(D(Y))s_{(1)}^{n+p-2}(Y) \cdots s_{(p)}^{n+1-(p+1)}(Y)$$

$$\cdot \int_{\mathscr{C}} \Pi f\left[\mathscr{C}^{-1}\begin{bmatrix} y_{1i} - \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{pi} - \mu_p \end{bmatrix}\right] \frac{d\mathscr{C}}{\sigma_{(1)}^{n+p} \cdots \sigma_{(p)}^{n+1}} \cdot d\boldsymbol{\mu}$$

$$g_S^*(\mathscr{C} : Y)\,d\mathscr{C} = k(D(Y))s_{(1)}^{n+p-2}(Y) \cdots s_{(p)}^{(n+1)-(p+1)}(Y)$$

$$\cdot \int_{R^p} \Pi f\left[\mathscr{C}^{-1}\begin{bmatrix} y_{1i} - \mu_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{pi} - \mu_p \end{bmatrix}\right] d\boldsymbol{\mu} \cdot \frac{d\mathscr{C}}{\sigma_{(1)}^{n+p} \cdots \sigma_{(p)}^{n+1}}.$$

27. The marginal structural distributions are

$$g_L^*(\mu : Y)\, d\mu = \frac{A_{n-p}}{A_n} \frac{|T(Y)|^{n-1}\, |T(Y)|_\nabla |T_\mu(Y)|_\Delta}{|T_\mu(Y)|^n\, |T_\mu(Y)|_\nabla\, |T(Y)|_\Delta} \Pi\, d\sqrt{n}\mu_j,$$

$$g^*(\mathcal{C} : Y)\, d\mathcal{C} = \frac{A_{n-1} \cdots A_{n-p}}{(2\pi)^{(n-1)p/2}} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} S(Y)\}$$

$$s_{(1)}^{n+p-2}(Y) \cdots s_{(p)}^{n-p}(Y) \frac{d\mathcal{C}}{\sigma_{(1)}^{n+p-1} \cdots \sigma_{(p)}^{n}},$$

where $T_\mu(Y)$ is the positive lower triangular square root of

$$S(Y) + n(\mathrm{m}(Y) - \mu)(\mathrm{m}(Y) - \mu)'.$$

29. (ii) Let $Y = [Y]\, D(Y)$ be the positive lower triangular and orthogonal factorization of Y:

$$[Y] = \begin{bmatrix} s_{(1)}(Y) & & & 0 \\ t_{21}(Y) & s_{(2)}(Y) & & \\ \cdot & \cdot & \cdot & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ t_{p1}(Y) & \cdots & t_{pp-1}(Y) & s_{(p)}(Y) \end{bmatrix}, \quad D(Y) = \begin{bmatrix} \mathrm{d}_1'(Y) \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{d}_p'(Y) \end{bmatrix};$$

$t_{j1}(Y), \ldots, t_{jj-1}(Y)$ and $s_{(j)}(Y)$ are the regression coefficients and residual length of y_j on $\mathrm{d}_1(Y), \ldots, \mathrm{d}_{j-1}(Y)$; $\mathrm{d}_j(Y)$ is the unit residual vector.

$$dm(Y) = \frac{dY}{|[Y]|^n}, \qquad d\mu(g) = \frac{dg}{|g|_\Delta}, \qquad d\nu(g) = \frac{dg}{|g|_\nabla}.$$

(iii) $dg^{-1} = \dfrac{dg}{|g|^{p+1}}$.

31. (i) $\dfrac{A_n \cdots A_{n-p+1}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2}(\Sigma s_{(j)}^2 + \Sigma t_{jj'}^2)\} s_{(1)}^{n-1} \cdots s_{(p)}^{n-p} \Pi\, ds_{(j)} \Pi\, dt_{jj'}$.

(ii) $\dfrac{A_n \cdots A_{n-p+1}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2} \operatorname{tr} S\} |S|^{(n-p-1)/2} \dfrac{dS}{2^p}$.

(iii) $$\frac{A_n \cdots A_{n-p+1}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} S(Y)\}$$

$$\cdot \left[\frac{s_{(1)}(Y) \cdots s_{(p)}(Y)}{\sigma_{(1)} \cdots \sigma_{(p)}}\right]^n \frac{s_{(1)}^p(Y) \cdots s_{(p)}^1(Y)}{s_{(1)}^1(Y) \cdots s_{(p)}^p(Y)} \frac{d\mathcal{C}}{\sigma_{(1)}^p \cdots \sigma_{(p)}^1}.$$

(iv) $$\frac{A_n \cdots A_{n-p+1}}{(2\pi)^{np/2}} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} S(Y)\}$$

$$\cdot \frac{|S(Y)|^{n/2}}{|\Sigma|^{n/2}} \frac{s_{(1)}^p(Y) \cdots s_{(p)}^1(Y)}{s_{(1)}^1(Y) \cdots s_{(p)}^p(Y)} \frac{d\Sigma}{2^p(\sigma_{(1)}^p \cdots \sigma_{(p)}^1)^2}.$$

36. (i) $$g_L(H : D)\, dH = k(D) \int_T f(T(HV + \underline{D})) \frac{|T|^n}{|T|_\Delta}\, dT \cdot dH,$$

$$g_L^*(\mathcal{B} : Y)\, d\mathcal{B} = \frac{k(D)}{|T(Y)|^r} \int_T f(TT^{-1}(Y)(\underline{Y} - \mathcal{B}V)) \frac{|T|^n}{|T|_\Delta}\, dT \cdot d\mathcal{B}.$$

(ii) $$g_S(T : D)\, dT = k(D) \int_B f(BV + T\underline{D})\, dB \cdot \frac{|T|^{n-r}\, dT}{|T|_\Delta},$$

$$g_S^*(\mathcal{C} : Y)\, d\mathcal{C} = k(D) \int_B f(BV + \mathcal{C}^{-1}\, T(Y)\underline{D})\, dB \frac{|T(Y)|^{n-r}}{|\mathcal{C}|^{n-r}} \frac{|T(Y)|_\nabla}{|T(Y)|_\Delta} \frac{d\mathcal{C}}{|\mathcal{C}|_\nabla}.$$

37. $$g([E] : D)\, d[E] = \prod_{j=1}^{p} \left[\frac{|VV'|^{1/2}}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}\mathrm{b}_j' VV' \mathrm{b}_j\}\, d\mathrm{b}_j\right]$$

$$\cdot \frac{A_{n-r} \cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}} \exp\{-\tfrac{1}{2}(\Sigma s_{(j)}^2 + \Sigma t_{jj'}^2)\}\, s_{(1)}^{n-r-1} \cdots s_{(p)}^{n-r-p}\, \Pi\, ds_{(j)}\, \Pi\, dt_{jj'},$$

$$g^*(\theta : Y)\, d\theta = \frac{|VV'|^{p/2} |\Sigma|^{-r/2}}{(2\pi)^{pr/2}} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(B(Y) - \mathcal{B})\, VV'(B(Y) - \mathcal{B})'\}\, d\mathcal{B}$$

$$\cdot \frac{A_{n-r} \cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} S(Y)\}\, s_{(1)}^{n-r+p-1}(Y) \cdots s_{(p)}^{n-r+1-p}(Y) \frac{d\mathcal{C}}{\sigma_{(1)}^{n-r+p} \cdots \sigma_{(p)}^{n-r+1}},$$

$$g_L^*(\mathcal{B} : Y)\, d\mathcal{B} = \frac{A_{n-r} \cdots A_{n-r-p+1}}{A_n \cdots A_{n-p+1}} \frac{|T(Y)|^{n-r} |T(Y)|_\nabla}{|T(Y)|_\Delta}$$

$$\cdot \frac{|T_{\mathcal{B}}(Y)|_\Delta}{|T_{\mathcal{B}}(Y)|^n\, |T_{\mathcal{B}}(Y)|_\nabla} |VV'|^{r/2}\, d\mathcal{B},$$

$$g_S^*(\mathcal{C} : Y)\, d\mathcal{C} = \frac{A_{n-r} \cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} S(Y)\}$$

$$\cdot\, s_{(1)}^{n-r+p-1}(Y) \cdots s_{(p)}^{n-r+1-p}(Y) \frac{d\mathcal{C}}{\sigma_{(1)}^{n-r+p} \cdots \sigma_{(p)}^{n-r+1}},$$

where

$$B = \begin{bmatrix} \mathrm{b}_1' \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{b}_p' \end{bmatrix}$$

and $T_{\mathcal{B}}(Y)$ is the positive lower triangular square root of

$$S(Y) + (B(Y) - \mathcal{B})VV'(B(Y) - \mathcal{B})'.$$

CHAPTER FOUR

1. (i) The structural distribution for θ given β is

$$g^*(\theta : \mathrm{x} : \beta)\, d\theta = k_\beta(\mathrm{d}(\mathrm{x}))\, \Pi f(x_i - \theta : \beta)\, d\theta.$$

(ii) The marginal likelihood function is $R^+(\mathrm{d})/k_\beta(\mathrm{d})$.

(iii) The marginal likelihood function is $R^+(\mathrm{d}) \exp\{-\Sigma d_i^2/2\sigma^2\}/\sigma^{n-1}$.

4. (ii) $$[X] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \bar{x}_1 & s(X) & & 0 \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \\ \bar{x}_p & 0 & & s(X) \end{bmatrix}, \quad D(X) = \begin{bmatrix} \dfrac{x_{11} - \bar{x}_1}{s(X)} & \cdots & \dfrac{x_{1n} - \bar{x}_1}{s(X)} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \dfrac{x_{p1} - \bar{x}_p}{s(X)} & \cdots & \dfrac{x_{pn} - \bar{x}_p}{s(X)} \end{bmatrix}.$$

(iii) $dm(X) = \dfrac{dX}{s^{np}(X)}, \; d\mu(g) = \dfrac{da\,dc}{c^{p+1}}, \; d\nu(g) = \dfrac{da\,dc}{c}, \; \Delta(g) = \dfrac{1}{c^p}.$

(iv) $g([E]:D:\beta)\,d[E] = k_\beta(D)\,\Pi f(\bar{e}_1 + sd_{1i}, \ldots, \bar{e}_p + sd_{pi}:\beta)s^{np-p-1}\,\Pi\, d\bar{e}_j\, ds,$

$$g^*(\theta:X:\beta)\,d\theta = k_\beta(D)\,\Pi f\left[\frac{x_{1i}-\mu_1}{\sigma}, \ldots, \frac{x_{pi}-\mu_p}{\sigma}:\beta\right]\frac{s^{np}(X)}{\sigma^{np}}\,\frac{1}{s^p(X)}\,\frac{d\boldsymbol{\mu}\,d\sigma}{\sigma}.$$

(v) The marginal likelihood is $R^+(D)/k_\beta(D)$.

5. (i) $\Pi\left(\left[\dfrac{n}{2\pi}\right]^{1/2}\exp\left\{-\dfrac{n\bar{e}_j^2}{2}\right\}d\bar{e}_j\right)\cdot\dfrac{1}{\Gamma((np-p)/2)}\exp\left\{-\dfrac{s^2}{2}\right\}\left[\dfrac{s^2}{2}\right]^{(np-p)/2-1}\dfrac{ds^2}{2}.$

(ii) $\Pi\left(\left[\dfrac{n}{2\pi\sigma^2}\right]^{1/2}\exp\left\{-\dfrac{n(\bar{x}_j-\mu_j)^2}{2\sigma^2}\right\}d\mu_j\right)\cdot\dfrac{1}{\Gamma[(np-p)/2]}$

$$\exp\left\{-\frac{s^2(X)}{2\sigma^2}\right\}\left[\frac{s^2(X)}{2\sigma^2}\right]^{(np-p)/2}\frac{d\sigma^2}{\sigma^2}.$$

10. (i) The structural distribution for θ given λ is

$$k(\mathbf{d}^\lambda)\,\Pi f(x_i^\lambda - \theta)\,d\theta.$$

(ii) The marginal likelihood for λ is

$$\frac{1}{k(\mathbf{d}^\lambda)}\,\Pi\left|\frac{dx_i^\lambda}{dx_i}\right|\left[\Sigma\left(\frac{dx_i^\lambda}{dx_i}\right)^{-2}\right]^{1/2}.$$

(iii) The marginal likelihood for λ (normal error) is

$$\frac{R^+(\mathbf{d}^\lambda)}{\sigma^{n-1}}\exp\left\{-\frac{\Sigma d_i^\lambda}{2\sigma^2}\right\}\Pi\left|\frac{dx_i^\lambda}{dx_i}\right|\left[\Sigma\left(\frac{dx_i^\lambda}{dx_i}\right)^{-2}\right]^{1/2}.$$

12. (i) The structural distribution for $\boldsymbol{\beta}$ given λ is

$$k(D_\lambda)\,\Pi f(y_i^\lambda - \Sigma\beta_u v_{ui})\,\Pi\, d\beta_u.$$

(ii) The marginal likelihood for λ is

$$\frac{R^+(D_\lambda)}{k(D_\lambda)}\,\frac{|J(\mathbf{y}:\lambda)|}{|VJ^{-2}(\mathbf{y}:\lambda)V'|^{-1/2}}.$$

13. (i) The structural distribution for $\boldsymbol{\beta}$ given λ is

$$\frac{|VV'|^{1/2}}{(2\pi\sigma_0^2)^{r/2}}\exp\left\{-\frac{1}{2\sigma_0^2}(\boldsymbol{\beta}-\mathbf{b}(\mathbf{y}^\lambda))'VV'(\boldsymbol{\beta}-\mathbf{b}(\mathbf{y}^\lambda))\right\}d\boldsymbol{\beta}.$$

(ii) The marginal likelihood function is

$$\frac{R^+(D_\lambda)}{\sigma_0^{n-r}}\exp\left\{-\frac{1}{2}\frac{\Sigma(d_i^\lambda)^2}{\sigma_0^2}\right\}\frac{|J(\mathbf{y}:\lambda)|}{|VJ^{-2}(\mathbf{y}:\lambda)V'|^{-1/2}}.$$

CHAPTER FIVE

1. (i) The distribution of error position is

$$\left[\frac{n}{2\pi\sigma^2}\right]^{1/2}\exp\left\{-\frac{n\bar{e}_1^2}{2\sigma^2}\right\}d\bar{e}_1\cdot\left[\frac{n}{2\pi\sigma^2}\right]^{1/2}\exp\left\{-\frac{n\bar{e}_2^2}{2\sigma^2}\right\}d\bar{e}_2\cdot\frac{dh}{2\pi}.$$

(ii) The structural distribution is

$$\left[\frac{n}{2\pi\sigma^2}\right]^{1/2}\exp\left\{-\frac{n}{2\sigma^2}(\bar{x}_1-\mu_1)^2\right\}d\mu_1\cdot\left[\frac{n}{2\pi\sigma^2}\right]^{1/2}\exp\left\{-\frac{n}{2\sigma^2}(\bar{x}_2-\mu_2)^2\right\}d\mu_2\cdot\frac{d\varphi}{2\pi}.$$

The first two parts give the marginal distribution for (μ_1, μ_2).

6. (ii) The indicated transformation variable is

$$[X] = \begin{bmatrix}1 & 0\\ \mathbf{x} & s(X)\,O(X)\end{bmatrix}$$

where $O(X)$ is the rotation matrix used in Problem 4.

(iii) $dm = dX/s^{np}(X)$; $d\mu = dg/c^{p+1}$; $d\nu = dg/c$; $\Delta = 1/c^p$.

(iv) $g([E]:D:\beta)\,d[E] = k_\beta(D)\,f([E]D:\beta)s^{np-p-1}\,d[E]$

$$g^*(\theta:X:\beta)\,d\theta = k_\beta(D)\,f(\theta^{-1}X:\beta)\frac{s^{np-p}(X)}{\sigma^{np+1}}\,d\boldsymbol{\mu}\,d\sigma\,d\Omega.$$

(v) $g^*(\boldsymbol{\mu}, \sigma:X:\beta)\,d\boldsymbol{\mu}\,d\sigma = k_\beta'(D)\,\Pi f\left[\dfrac{x_{1i}-\mu_1}{\sigma}, \ldots, \dfrac{x_{pi}-\mu_p}{\sigma}:\beta\right]\dfrac{s^{np-p}(X)}{\sigma^{np+1}}\,d\boldsymbol{\mu}\,d\sigma.$

(vi) The marginal likelihood function is $R^+(D)/k_\beta(D)$.

7. (i) The error position distribution is

$$g([E]:D)\,d[E] = \prod_1^p\left(\sqrt{\frac{n}{2\pi}}\exp\left\{-\frac{n\bar{e}_j^2}{2}\right\}d\bar{e}_j\right)$$

$$\cdot\frac{A_{np-p}}{(2\pi)^{(n-1)p/2}}\exp\left\{-\frac{s^2}{2}\right\}s^{np-p-1}\,ds\cdot\frac{do_1\cdots do_{p-1}}{A_p\cdots A_2}.$$

(ii) The structural distribution is

$$g^*(\theta:X)\,d\theta = \prod_1^p\left(\left[\frac{n}{2\pi\sigma^2}\right]^{1/2}\exp\left\{-\frac{n(\bar{x}_j-\mu_j)^2}{2\sigma^2}\right\}d\mu_j\right)$$

$$\cdot\frac{A_{np-p}}{(2\pi)^{(n-1)p/2}}\exp\left\{-\frac{s^2(X)}{2\sigma^2}\right\}\left[\frac{s(X)}{\sigma}\right]^{np-p}\frac{d\sigma}{\sigma}\cdot\frac{d\omega_1\cdots d\omega_{p-1}}{A_p\cdots A_2}.$$

The structural distribution for location is

$$g_L^*(\boldsymbol{\mu}:X)\,d\boldsymbol{\mu} = \frac{A_{np-p}}{A_{np}}\,\frac{n^{p/2}s^{-p}(X)}{\left[1+\Sigma\dfrac{n(\mu_j-\bar{x}_j)^2}{s^2(X)}\right]^{np}}\,\Pi\,d\mu_j.$$

9. (i) G is unitary on R^{pn} provided points Y having $\mathbf{y}_1, \ldots, \mathbf{y}_p$ linearly dependent are excluded.

(ii) Let the rows of $D(Y)$ be an orthonormal basis correctly oriented for $L^+(\mathbf{y}_1, \ldots, \mathbf{y}_p)$; let $C(Y)$ be the matrix that produces the basis $\mathbf{y}_1, \ldots, \mathbf{y}_p$ from the basis $\mathbf{d}_1, \ldots, \mathbf{d}_p$.

(iii) $dm = \dfrac{dY}{|C(Y)|^n}$; $\quad d\mu = \dfrac{dg}{|g|^p}$; $\quad d\nu = \dfrac{dg}{|g|^p}$; $\quad \Delta = 1.$

(iv) $\left|\dfrac{\partial g^{-1}}{\partial g}\right| = |g|^{-2p}.$

10. (i) The distribution for error is

$$g(C:D)\,dC = k(D)\,\Pi f\left(C\begin{pmatrix} d_{1i} \\ \cdot \\ \cdot \\ \cdot \\ d_{pi}\end{pmatrix}\right)|C|^{n-p}\,dC.$$

(ii) The structural distribution is

$$g^*(\Gamma:Y)\,d\Gamma = k(D(Y))\,\Pi f\left(\Gamma^{-1}\begin{pmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi}\end{pmatrix}\right)\frac{|C(Y)|^n}{|\Gamma|^n}\frac{d\Gamma}{|\Gamma|^p}.$$

11. (i) The distribution for the quantity $\mathcal{T}$ is

$$k(D)\prod_2^p A_j\,\Pi f\left(\mathcal{T}^{-1}\begin{pmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi}\end{pmatrix}\right)\frac{|T(Y)|^n}{|\mathcal{T}|^n}\frac{d\mathcal{T}}{|\mathcal{T}|_\Delta}.$$

(ii) The structural distribution for Σ is

$$k(D)\prod_2^p A_j\,\Pi f\left(\mathcal{T}^{-1}\begin{pmatrix} y_{1i} \\ \cdot \\ \cdot \\ \cdot \\ y_{pi}\end{pmatrix}\right)\frac{|T(Y)|^n}{|\Sigma|^{(n+p+1)/2}}\frac{d\Sigma}{2^p}.$$

12. (i) $\dfrac{A_n\cdots A_{n-p+1}}{(2\pi)^{np/2}}\exp\{-\tfrac{1}{2}(\Sigma s_{(j)}^2 + \Sigma t_{jj'}^2)\}\,s_{(1)}^{n-1}\cdots s_{(p)}^{n-p}\,\Pi\,ds_{(j)}\,\Pi\,dt_{jj'}\cdot\dfrac{dO}{A_p\cdots A_2}.$

(ii) $\dfrac{A_n\cdots A_{n-p+1}}{(2\pi)^{np/2}}\exp\{-\tfrac{1}{2}(\Sigma s_{(j)}^2 + \Sigma t_{jj'}^2)\}\,s_{(1)}^{n-1}\cdots s_{(p)}^{n-p}\,\Pi\,ds_{(j)}\,\Pi\,dt_{jj'}.$

(iii) $\dfrac{A_n\cdots A_{n-p+1}}{(2\pi)^{np/2}}\exp\{-\tfrac{1}{2}\operatorname{tr} S\}\,|S|^{(n-p-1)/2}\,\dfrac{dS}{2^p}.$

(iv) $\dfrac{A_n\cdots A_{n-p+1}}{(2\pi)^{np/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\dfrac{|S(Y)|^{n/2}}{|\Sigma|^{n/2}}\dfrac{d\mathcal{T}}{|\mathcal{T}|_\Delta}\cdot\dfrac{d\Omega}{A_p\cdots A_2}.$

(v) $\dfrac{A_n\cdots A_{n-p+1}}{(2\pi)^{np/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\dfrac{|S(Y)|^{n/2}}{|\Sigma|^{n/2}}\dfrac{d\mathcal{T}}{|\mathcal{T}|_\Delta}.$

(vi) $\dfrac{A_n\cdots A_{n-p+1}}{(2\pi)^{np/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\dfrac{|S(Y)|^{n/2}}{|\Sigma|^{(n+p+1)/2}}\dfrac{d\Sigma}{2^p}.$

20. (i) $g_L(H:D)\,dH = k(D)\displaystyle\int_C f(C(HV+\underline{D}))\,|C|^{n-p}\,dC\cdot dH,$

$$g_L^*(\mathcal{B}:Y)\,d\mathcal{B} = \frac{(k(D)}{|C(Y)|^r}\int_C f(CC^{-1}(Y)(\underline{Y}-\mathcal{B}V))\,|C|^{n-p}\,dC\cdot d\mathcal{B}.$$

(ii) $g_S(C:D)\,dC = k(D)\displaystyle\int_B f(BV+C\underline{D})\,dB\cdot|C|^{n-p-r}\,dC,$

$$g_S^*(\Gamma:Y)\,d\Gamma = k(D)\int_B f(BV+\Gamma^{-1}C(Y)\underline{D})\,dB\cdot\frac{|C(Y)|^{n-r}}{|\Gamma|^{n-r}}\frac{d\Gamma}{|\Gamma|^p}.$$

21. (i) $k(D)\displaystyle\prod_2^p A_j f(\mathcal{T}^{-1}(\underline{Y}-\mathcal{B}V))\,\dfrac{|C(Y)|^{n-r}}{|\mathcal{T}|^n}\dfrac{d\mathcal{B}\,d\mathcal{T}}{|\mathcal{T}|_\Delta}.$

(ii) $k(D)\displaystyle\prod_1^p A_j f(\mathcal{T}^{-1}(\underline{Y}-\mathcal{B}V))\,\dfrac{|C(Y)|^{n-r}}{|\mathcal{T}|^{n+p+1}}\dfrac{d\mathcal{B}\,d\Sigma}{2^p}.$

22. (i) $\dfrac{|VV'|^{p/2}}{(2\pi)^{rp/2}}\exp\{-\tfrac{1}{2}\displaystyle\sum_1^p \mathbf{b}_j'VV'\mathbf{b}_j\}\prod_1^p d\mathbf{b}_j\cdot\dfrac{A_{n-r}\cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}}\exp\{-\tfrac{1}{2}(\Sigma s_{(j)}^2+\Sigma t_{jj'}^2)\}$

$$s_{(1)}^{n-r-1}\cdots s_{(p)}^{n-r-p}\,\Pi\,ds_{(j)}\,\Pi\,dt_{jj'}\cdot\frac{dO}{A_p\cdots A_2}.$$

The marginal distribution of (B, T) is given by the first two parts of the preceding distribution.

(ii) $\dfrac{A_{n-r}\cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr} S\}\,|S|^{(n-r-p-1)/2}\,\dfrac{dS}{2^p}.$

(iii) $\dfrac{|VV'|^{p/2}\,|\Sigma|^{-r/2}}{(2\pi)^{rp/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}(B(Y)-\mathcal{B})VV'(B(Y)-\mathcal{B})'\}\,d\mathcal{B}$

$$\cdot\frac{A_{n-r}\cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\frac{|S(Y)|^{(n-r)/2}}{|\Sigma|^{(n-r)/2}}\frac{d\mathcal{T}}{|\mathcal{T}|_\Delta}\frac{d\Omega}{A_p\cdots A_2}.$$

(iv) $\dfrac{A_{n-r}\cdots A_{n-r-p+1}}{A_n\cdots A_{n-p+1}}\dfrac{|VV'|^{p/2}\,|S(Y)|^{-r/2}}{|I+S^{-1}(Y)(B(Y)-\mathcal{B})VV'(B(Y)-\mathcal{B})'|^{n/2}}\,d\mathcal{B};$

$$\frac{A_{n-r}\cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\frac{|S(Y)|^{(n-r)/2}}{|\Sigma|^{(n-r)/2}}\frac{d\mathcal{T}}{|\mathcal{T}|_\Delta}.$$

(v) $\dfrac{A_{n-r}\cdots A_{n-r-p+1}}{(2\pi)^{(n-r)p/2}}\exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}S(Y)\}\dfrac{|S(Y)|^{(n-r)/2}}{|\Sigma|^{(n-r+p+1)/2}}\dfrac{d\Sigma}{2^p}.$

23. The simplified matrix-variate t-distribution:

$$\frac{A_{n-r}\cdots A_{n-r-p+1}}{A_n\cdots A_{n-p+1}}\frac{|VV'|^{p/2}}{|I+HVV'H'|^{n/2}}\,dH.$$

CHAPTER SIX

2. $l(x,\theta_0) = -\theta_0\ln(\ln x)$. Linearization with respect to $\theta^* = -\ln\theta$ gives $l^* = \ln\ln x$ with distribution function $1-\exp\{-\exp\{l^*-\theta^*\}\}$.

CHAPTER SEVEN

3. The squared length of the difference vector is 15.948037; $\chi^2 = 63.792$. The 1 and $\frac{1}{2}$% points for χ^2 on 5 are 15.086 and 16.750; there is strong evidence against the hypothesis of symmetry.

Index